BerkeleyHaas
Haas School of Business
University of California Berkeley

Letting the Computers Take Over:

Using AI to Solve Marketing Problems

California Management Review
1–30

DOI: 10.1177/0008125619859318
journals.sagepub.com/home/cmr
SAGE

Gijs Overgoor[1,2], Manuel Chica[3,4], William Rand[2], and Anthony Weishampel[2]

SUMMARY

Artificial intelligence (AI) has proven to be useful in many applications from automating cars to providing customer service responses. However, though many firms want to take advantage of AI to improve marketing, they lack a process by which to execute a Marketing AI project. This article discusses the use of AI to provide support for marketing decisions. Based on the established Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, it creates a process for managers to use when executing a Marketing AI project and discusses issues that might arise. It explores how this framework was used to develop three cutting-edge Marketing AI applications.

KEYWORDS: artificial intelligence, analytics, marketing, social networks, social media

Artificial intelligence (AI) is one of the most popular buzzwords in business today but that is for a very good reason: AI has shown to be a very powerful tool for many marketing applications. AI has been around for decades, but its recent popularity is due to three major factors: the growth of Big Data; the availability of cheap, scalable, computational power; and the development of new AI techniques. In the past, one of the problems with many AI methods was that they required a lot of data in order to train, but before the advent of the Big Data revolution, those data were often hard to come by. Moreover, even when large-scale data were available, it

[1]University of Amsterdam, Amsterdam, The Netherlands
[2]North Carolina State University, Raleigh, NC, USA
[3]University of Granada, Granada, Spain
[4]The University of Newcastle, Newcastle, Australia

would often take way too long to actually train AI models on these data. The development of new high-performance computing systems that can parallelize this process has made that cheaper and faster than ever before. Finally, new AI methods, such as deep learning, have been developed that can take advantage of both large-scale data and cheap computational power at the same time.[1]

We have already seen the potential impact of AI on marketing, as illustrated by the power of Amazon's recommender systems for product purchases and "anticipatory" one-hour shipping or by Google's ability to automatically pair advertising with content.[2] In the near future, AI is expected to make marketing more efficient by speeding up the decision-making process and providing marketing managers with information and insights that they could not develop in any other way.

There has been good academic research into examples of how AI can facilitate marketing. For instance, AI has been shown to help out marketing by the use of text mining to help understand online word-of-mouth (WOM),[3] modeling direct marketing responses using evolutionary programming,[4] predicting churn using classification trees,[5] and adapting websites automatically to better fulfill customer needs,[6] among many other applications. Yet, there is still a need for more research into how AI can help solve marketing problems.[7] For Marketing AI to be truly successful, managers need to be better equipped to understand how to implement a Marketing AI solution, and that is one area that has not been well researched yet.

This article presents a framework, based on the popular Cross-Industry Standard Process for Data Mining (CRISP-DM) framework,[8] that lays out the steps to take when using AI to help solve a marketing problem.

AI, Machine Learning (ML), Data Mining, and Analytics

One classic textbook in the field by Russell and Norvig defines *artificial intelligence* as the study of the "general principles of rational agents and on components for constructing them."[9] "Agents" in this context refers to any system that can perceive the world around it in some way and take action on the basis of those perceptions. A rational agent, in Russell and Norvig's explanation, is "one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome."[10] Russell and Norvig deliberately steer away from using humans as the measure for AI since that can be very hard to define, while rationality is much easier to assess scientifically. Moreover, in many ways, this definition is more useful from a marketing perspective, since it emphasizes making the best decision possible with the given information. *Marketing AI* can now be defined as the development of artificial agents that, given the information they have about consumers, competitors, and the focal company, suggest and/or take marketing actions to achieve the best marketing outcome. Some examples of Marketing AI that meet this definition would include chatbots for customer service, tools that model the potential outcomes of a new marketing campaign, recommender systems that help managers choose content for online marketing, or models that identify latent characteristics of consumers that are predictive of future interactions with the company.

AI has recently become popular because it provides a cheap way to make predictions about complex problems based on examples in historical data that a company might already have. Machines are often able to predict better than humans and they can do it much faster. Especially with significant improvements in computational power and data availability in the last decade, the cost of prediction has dropped significantly, leading to a dramatic increase in the popularity of AI.[11] A formal definition of ML from one of the classic textbooks in the field states, "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*."[12] In other words, if a computer program can improve performance of a task based on an experience with that task and does so because of that experience, then it has learned. In practice, ML is "a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty."[13] Although AI is broader than just ML, most of the applications of Marketing AI are in the ML space.

Although they are often used together, AI and ML differ from the "classical" modeling approach, often statistical modeling, that is traditionally used in Marketing in several ways. First, ML and classical models have different goals. Classical models emphasize causal and explanatory relationships, whereas ML focuses on operational effectiveness and prediction accuracy.[14] Second, classical models require the modeler to understand the relation and implementation that the variable has on an equation in an effort to best estimate the function output to a certain error, generally based on theory. Econometrics, for example, is defined as "the interaction of economic theory, observed data and statistical methods."[15] In contrast, besides providing a cheap method for prediction, ML can learn from data without relying on assumptions or rules-based programming and it can often model much more complex interactions between variables.[16] Marketing AI solutions are often a combination of AI and classical methods, where AI is used to make predictions or automate processes and traditional methods are used to create an understanding of underlying relationships and mechanisms. For example, in many cases, ML is used to make predictions about complex data sources and this information is then used as input for an econometric model.

Another term that is often used in related projects is *data mining*, and this term is important for this discussion because the framework that we are going to use to develop our Marketing AI process approach is actually drawn from data mining. Data mining is sometimes referred to as knowledge discovery from databases, and it is the study of identifying patterns in data.[17] To this extent, data mining is often part of ML since ML needs the patterns identified by data mining in order to create rules to predict future behavior. One particularly exciting example of data mining in marketing is Amazon's retail forecasting methods where they use AI to anticipate product demand far enough in advance that they can make sure the product is stored nearby before the consumer decides to buy, allowing them to ensure one-hour delivery of their products.[18] This kind of customer tracking is one example of discovering purchase patterns through data mining.

Finally, how does this all relate to marketing analytics? Shmueli et al. define *business analytics* as "the practice and art of bringing quantitative data to bear on decision making."[19] So *marketing analytics* is bringing quantitative data to bear on marketing decision making. AI, ML, and data mining are all techniques that can help in making better decisions using data, but there are other related methods, such as classical models that could also be used to develop marketing analytics; thus, though these terms, artificial intelligence, machine learning, data mining, and marketing analytics, all overlap and relate to each other, they are distinct in their own right.

Methods and Toolkits

Machines are now better, cheaper, and faster at making predictions and there is more data available than ever. Generally, there are three ways a machine learns.

Supervised learning. In this setup, there are data with labeled responses available and the machine learns how to recognize the labels based on the data. The data are split up into training and testing data. The machine learns on the training data and is then evaluated by comparing the predicted labels and the true labels. After the machine is trained and evaluated, it can then be deployed and predict the labels based on the new examples that it has not seen before, for example, if you know the historical customer lifetime value (CLV) of a group of customers and a set of characteristics about those customers. The machine can learn a model that relates the characteristics of the customers to CLV. Then, in test time, the machine is presented with customer characteristics and predicts what the CLV of the customer will be. The effectiveness of the model is measured by how well it does on the testing data. Essentially, supervised learning is where you teach the machine by showing examples. Examples of supervised learning include classification, support vector machine, and decision trees.

Unsupervised learning. In these problems, there is unlabeled data available and the machine has no information on what the data represent. The machine will then learn to recognize patterns and similarities in the data. Consequently, it can group certain observations or recognize patterns. In the case of unsupervised learning, the machine learns without a teacher. For example, if a firm wanted to cluster their customers, then unsupervised learning could be used to automatically identify the clusters that have the most in common. A manager would still have to figure out what the clusters represent, but this approach can be quite powerful. Examples of unsupervised learning include clustering and anomaly detection.

Reinforcement learning. Reinforcement learning is similar to unsupervised learning except now the machine learns by getting some feedback after taking actions.

The machine takes actions based on a predicted reward structure and learns by adjusting strategy based on the difference between the predicted and realized output of an action. Reinforcement learning means the machine learns by trial and error, but the trials that it attempts are guided by the model and the algorithm. An example would be trying to learn what order of ads to show to a customer in order to encourage them to make a purchase. Eventually, at each time point, the model has to make a decision about what ad to show based on the interest the consumer has shown so far, but it does not know if it has made the right decision until the customer makes an actual purchase. Examples include Q-learning and adversarial networks.

ML methods do not necessarily only learn in one of these three ways. It is possible for an ML method to be used for supervised, unsupervised, and reinforcement learning. For instance, a method in ML that has recently become very popular due to the increase of data availability and computer power is deep learning. Deep learning methods have shown to work well for all three categories of ML. There are various deep neural network structures that apply to different types of data that work particularly well because of the architecture of the network. For example, convolutional neural networks (CNNs) work well for image classification, whereas recurrent neural networks are used for sequence models, such as time series.[20]

Another important aspect of ML is whether the learning occurs online or offline. The main difference between online and offline ML is that with online ML the model learns based on one incoming observation at a time whereas offline learning uses all the available data at once to learn a model across all of the data. Overall, online ML can be faster and more efficient; however, the accuracy compared with offline learning is often lower.[21] For example, it is possible to do online deep learning, but deep learning generally requires a lot of offline training time and data to get the desired accuracy. The examples presented in this research mostly concern the use of offline ML. The machines learn offline (i.e., they are pre-trained on offline data) and are then deployed to make predictions on new data. Even though they make predictions on newly incoming online data, the learning has occurred offline. Ideally, the end result of the examples discussed will be that after deployment, the machines are able to learn online and in real time as well, or the offline models are updated on a regular basis using the new data. The different categories of learning can occur online as well as offline, but some categories lend themselves better for offline learning, such as supervised learning, and others generally work well online, such as reinforcement learning.

Complementary to ML, statistical models (described earlier as one of the "classical" models) play an important role in Marketing AI projects. Statistical models require a more theoretical understanding and intuition behind methods, the variables, and their relationship to the business problem. Examples of statistical models not only include econometric methods as simple as regression analysis but also include methods such as support vector machines and causal state modeling. Regression analysis is a group of mathematical procedures for studying the relationship between an outcome variable and a set of explanatory variables.[22] Support vector machines are statistical models that are used to analyze data for classification and also regression (i.e., support vector regression).[23] Support vector machines are an example of a statistical model that uses ML to classify by optimizing their parameters to allow the maximum separation between the classes, being also able to solve regression problems.

In addition to statistical models and ML, another tool that is commonly used as a component of AI is computational modeling, often agent-based modeling (ABM) in particular. ABM is a model where rules of behavior are written for autonomous agents that have their own properties and behaviors, and then those agents are embedded in a computational environment where they can interact.[24] ABM by itself is not AI. ABM is a modeling framework, just as game theory is a modeling framework. ABM is a way that you can describe a system but does not necessarily need to use any form of AI. However, ABM and other computational modeling frameworks often use AI. ML, for example, can be used to optimize and calibrate agent actions and interactions with each other and the system in an agent-based model, which we will explore in one of the case studies below. Because both computational models and AI are written in computer code and often used in Marketing to understand human behavior, they are sometimes conflated. They are distinct, but related methods.

In the end, building any Marketing AI requires a great understanding of the business, the data, and the methods, which makes human judgment an ever important part of AI.[25] Judgment is necessary to determine the trade-offs of certain actions and to determine what makes an accurate prediction. Our framework provides the right questions to ask at the different stages of a Marketing AI project for effective judgment.

There are three ways AI can be implemented:

- Writing code from scratch, using a programming language that works well for the task at hand. Popular examples are R and Python.
- Using prebuilt packages or libraries in programming languages such as R, Python, MATLAB, and SAS. This is similar to writing from scratch, except you make use of pre-coded functions and scripts for a family of AI methods that have been previously written and distributed. Examples of these packages are as follows:
    - Scikit-learn: An ML package for Python that has several methods and libraries built-in such as classification, regression, clustering, model selection, and data preprocessing.[26]
    - Mlr: The scikit-learn equivalent for R.[27]
    - Rpart: An R package for recursive partitioning and regression trees.[28]
    - Dplyr: An R package focused on data wrangling. This is mostly used for data processing and structuring.[29]

  - TensorFlow: An open-source ML and deep learning framework that runs on Python.[30]
  - Keras: A high-level neural network API (application programming interface), written in Python and capable of running on top of TensorFlow. Keras makes it easy to build deep neural network architectures by using pre-trained models or standardized layers of the networks.[31]
  - SciPy: A library that combines several AI-related packages for Python. It includes an interactive console called iPython.[32]
  - ML Toolbox: A toolbox for ML in MATLAB.
  - SAS: SAS, one of the main business analytics frameworks in the industry, has ML and AI-related methods built-in.
  - Tableau: A visualization software that is very convenient for making raw data understandable and visual.
- Using "plug and play" software that provides a user-friendly tool to implement the methods described:
  - Weka and Orange: These open-source software tools, both developed in Java, are two of the standard tools in academia for ML and data mining. They have a comprehensive set of algorithms for supervised and unsupervised data mining tasks as well as preprocessing, post-processing, and visualization. In addition, the user can use its visual work flow to design the experiments and manage their own datasets. An API can be used to link the algorithms from the user's own code.[33]
  - Knowledge Extraction based on Evolutionary Learning (KEEL): KEEL is an open-source platform developed in Java that makes use of AI methods (i.e., evolutionary algorithms and fuzzy set theory) to evolve ML algorithms such as classification, regression, clustering, or association rules.[34]

In general, coding your own models and tools from scratch allows you the most flexibility for tailoring your model, while using the plug and play approach gives you the least flexibility, with the prebuilt tool kits approach being somewhere in the middle. However, the trade-off is development time. The plug and play tools can be used on a new dataset in very little time, while writing your own code from scratch can take a considerable amount of time. There are many additional tools that are available for the implementation and creation of Marketing AI tools, but this list is a good introduction to many of the most powerful tools.

The CRISP-DM Framework

The CRISP-DM[35] was not originally developed for applying AI methods to business processes, but it provides a strong basis for such a framework. We adapt this framework to help in deciding when and how to use AI and ML to solve marketing problems. CRISP-DM was created by a consortium of companies working together in 1996, and though there are other frameworks for the development of data mining solutions,[36] CRISP-DM is the most widely taught and used.[37] The goal of the CRISP-DM project was to create an open process model to describe a standard approach to use during data mining and analytics. In this way, the CRISP-DM was envisioned to be a *best practices* model of how to conduct data mining work. CRISP-DM was developed as a hierarchical description, so every phase can be unpacked to additional phases and so on, all the way down to the actual implementation of the project, but here we only describe the high-level phases, the aspects that relate to those phases, and how they can be used for Marketing AI.[38]

CRISP-DM has been used before in many different contexts from manufacturing[39] to bank fraud,[40] and even in marketing contexts,[41] but to our knowledge, it has not been explored substantially in the academic marketing literature, despite calls that marketing researchers themselves employ a more rigorous method of data mining.[42] As part of this project, we had originally considered creating a new process model for Marketing AI, but the steps of CRISP-DM are well defined and well accepted. Here, however, we adapt the details and descriptions of the CRISP-DM framework to a Marketing AI context. The major phases of the CRISP-DM process are the following.

Business Understanding

The first goal when considering whether to employ Marketing AI in any context is to determine the marketing objectives. What is this marketing action or decision trying to achieve? For instance, in the image selection project for the online travel agency described below, the goal was to increase click-through rate (CTR) for hotel listings, but often the answer is to increase sales. At this point, it is often useful to assess the situation. What is currently being done to achieve the marketing goals? In many cases in marketing, the answer will be that humans are currently making the decisions that we either want the computer to make, or no one is making the decision in any structured way currently, and the computer can help to make decisions in this space. Once this has been done, then it is possible to determine the Marketing AI goals. How will the success of the project be determined? Once these questions have been answered, then it is possible to start to produce a Marketing AI project plan. This involves scoping out the rest of the steps described below.

Data Understanding

AI, in general, is highly reliant on data. In fact, there is some evidence in the AI world that data are more important than the model.[43] Regardless, understanding the data will be critical to any Marketing AI project. The first part of this phase will be to collect the initial data. Identify which data are relevant to the project and then describe the data in detail, preferably using a data dictionary, which is essentially a formalized description of all of the data that can be used to discuss the data among team members who may have different backgrounds. Once the data have been collected and described, then it is important to explore the data. Often this task is guided by the marketing objectives, so the process revolves around trying to identify what factors of the data are associated with the objectives. For instance, how many conversions are we making per day and what is the average value of those conversions? Usually, this exploration is best handled by visualizing the data to illustrate and explore patterns. At this time, it is also useful to verify data quality. This involves checking to make sure that there is not any missing data or that the data actually make sense.

Data Preparation

Data preparation is where most time is spent on a Marketing AI project; even more than on the modeling efforts. The first part of this phase is to select the data, which means choosing exactly which data need to be incorporated into the Marketing AI solution both for development and testing. The data will often need to be cleaned at this point. Cleaning involves making sure that all of the data look similar in structure. This could involve removing data, that is, missing values, or normalizing the data to enable easy comparison between different types of data. Besides cleaning the data, it may also be necessary to add to it. In some cases, the raw data are inappropriate for modeling and it is better to construct new data, which is often done by constructing derived values from the raw data, for instance, taking textual data and tokenizing it.[44] It may also be necessary to integrate data. For instance, if the data are spread over many files with different columns, it may be easier and may in fact provide new insights to bring all those data together into one table or repository. After those data have been properly integrated, the final step in data preparation involves formatting the data appropriately. If using an off-the-shelf AI tool, such as Keras, then it is often the case that the data need to be formatted in a specific way.[45]

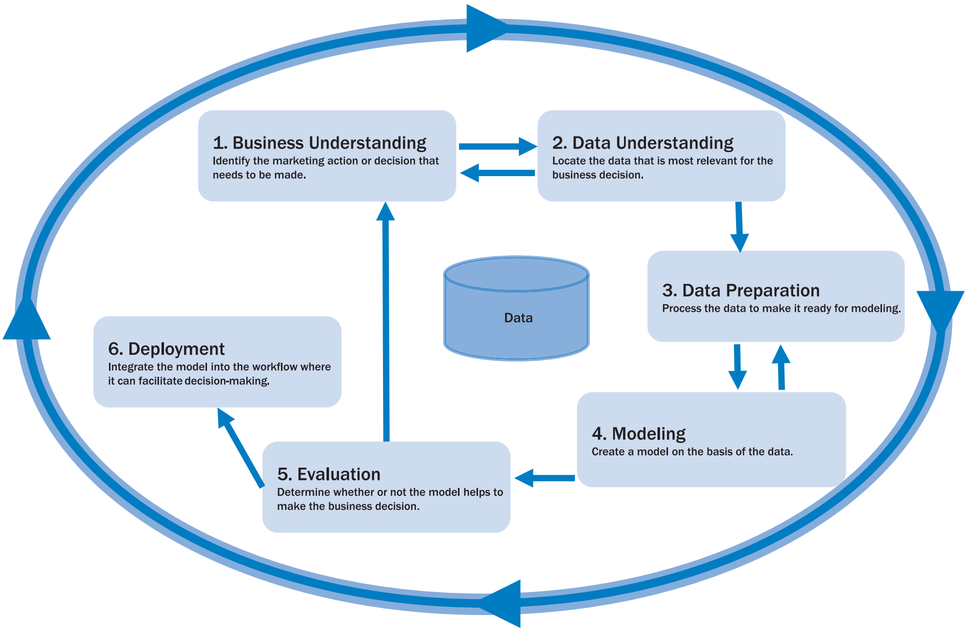
Modeling

When many people think of creating Marketing AI, this is the phase that they are actually thinking of. How do we build the model that will help us make a decision automatically? The first aspect will be to identify the modeling technique or techniques. This involves figuring out which approach—from neural nets to decision trees to ABM to linear regression—best solves the problem being examined. In some cases, the answer is to explore multiple modeling methods and then assess which one performed the best. Once the techniques have been chosen, the next step is to generate the testing criteria. How will the model be assessed? Often this involves taking a set of data and holding it out and identifying a metric of performance that will be used to assess the model. Unlike statistical models, it is often the case that Marketing AI models are built on one set of data and then tested on another. This additional dataset is sometimes called a hold-out set or a test set. The idea is that if the model generalizes from another dataset (often called the training dataset) to the testing dataset, then it is more likely that it will also turn out to be useful in data that have not been seen.[46] A dataset is either split into two datasets, a training and testing set, or into three datasets, a training, validation, and testing set. There is no general rule on the best way to split up the data because of several factors that influence the performance such as the sample size, signal-to-noise ratio, the number of hyperparameters, and the general complexity of a model.[47] We often see a split of about 50% to 90% to training data, and about 10% to 50% to testing or validation and testing data. Given enough data, one approach is to use a learning curve, which plots the size of the training set against the accuracy of the model to help determine a suitable training set size.[48] Often, we are searching for the optimal bias-variance composition.[49] Bias is an error measure of how much the model systematically differs from true results, while variance is an error measure of how different the model results are from each other. The bias-variance trade-off is about finding the right amount of model complexity to achieve the best prediction accuracy and to minimize these sources of error. At times, a model might be too simple when it has very few parameters and a low variance, that is, the predictions are all very similar, but the predictions are systematically off of the true value, that is, there is a high bias and the model underfits the data. A solution is to increase the complexity of the model, in which case the predictions might now be more accurate on average, however more spread out, that is, there is a higher variance and the model potentially overfits the data. Overfitting occurs because along with the underlying pattern, the model fits the noise, or outliers, in the data. An optimal bias-variance trade-off aims for a high prediction accuracy that does not overfit or underfit the data. Once the size of the verification and validation sets have been determined, the next step is to build the model using the training set and, in some cases, to fine-tune it on the validation set. After the model has been built, it can then be assessed by examining its performance on the testing set. This is the standard by which the Marketing AI solution will be assessed. It is important to note that this is an iterative process. After assessment, the model can be adjusted and re-trained for improvement and then re-assessed again until the results are satisfactory. It is also common for Marketing AI projects to go through the steps of data understanding, data preparation, and modeling multiple times, because sometimes ML can help to understand or prepare the data, even before a full model is built to help solve the business problem. For example, a first-level ML model can be pre-trained on offline data, sometimes unrelated to the project, for it to learn to recognize patterns or make predictions about the complex data formats that the final model will be examining. Subsequently, this pre-trained model is then used to prepare data or make predictions on new data that will be used in a second-level model that helps to solve the business problem. For instance, a neural network could be used for identifying concepts in images, and then those concepts can be fed to another model that relates those concepts to some final outcome, such as engagement with the image.

Evaluation

Now that we have a model and we have assessed its performance, it is time to evaluate the results. Have we in fact met the goals that we laid out in the first step of this project? As part of this phase, it is often useful to review the process that was used to arrive at this model and determine whether all the data are still available and can be made available in a way to facilitate deployment of the model; finally, determine next steps. Is there more to be done or is this it? If we have met our goals, then it is time to move on to actually deploying the solution.

FIGURE 1. Illustration of the CRISP-DM process.



Note: CRISP-DM = Cross-Industry Standard Process for Data Mining.

Deployment

The final step is to deploy the Marketing AI solution in a way that will actually increase business value. As part of this, it is necessary to plan deployment to understand exactly when and how the tool will be implemented. An important aspect of any major change is also to plan how to monitor and maintain the tool. If this process has been carried out correctly, then the tool should be well designed at this stage, but it may become less accurate over time and that needs to be monitored and assessed on a regular basis. A final report should also be put together and the whole process should be reviewed and takeaways for future similar projects should be discussed and actions should be taken to make sure the process improves every time. One of the things that is very different about Marketing AI, as opposed to traditional data mining, is that when deployed, a Marketing AI can be set up to continually update itself using new data. In the example below of the WOM decision support systems (DSS), the simulation model could be improved automatically after each marketing campaign to reflect the newest results; while in the example of image selection, the Marketing AI could continually update what aspects of an image are important.

In a common visualization of CRISP-DM, these phases are featured as flowing into each other, but as you can see from Figure 1, there are some backward arrows as well. For instance, early on, it is often necessary to iterate back and forth between business understanding and data understanding. As more (or less) data are found than the modelers had anticipated, then it may be useful to revise the scope of the project up or down. Another backward arrow exists between data preparation and modeling, since it may be necessary to extract additional features from the data to facilitate modeling. Finally, once the model has been built and is being evaluated, it is necessary to make sure it fulfills the goals laid out in the business understanding phase. In many cases, even after the whole process is finished, that is not the end. Instead, the whole process will start back over with a new Marketing AI problem, building on the learning from the last project.

These arrows indicate the standard flow for the CRISP-DM process, but at times it is necessary to move back and forth between different stages for other reasons. We have highlighted a few of those in the examples below and illustrated them in the accompanying diagrams with dashed arrows.

The Examples

Now that we have described how to use the CRISP-DM framework to implement a Marketing AI project at an abstract level, we provide some examples of how this process can be carried out using three distinct, real projects that are on the leading edge of AI applications in marketing.

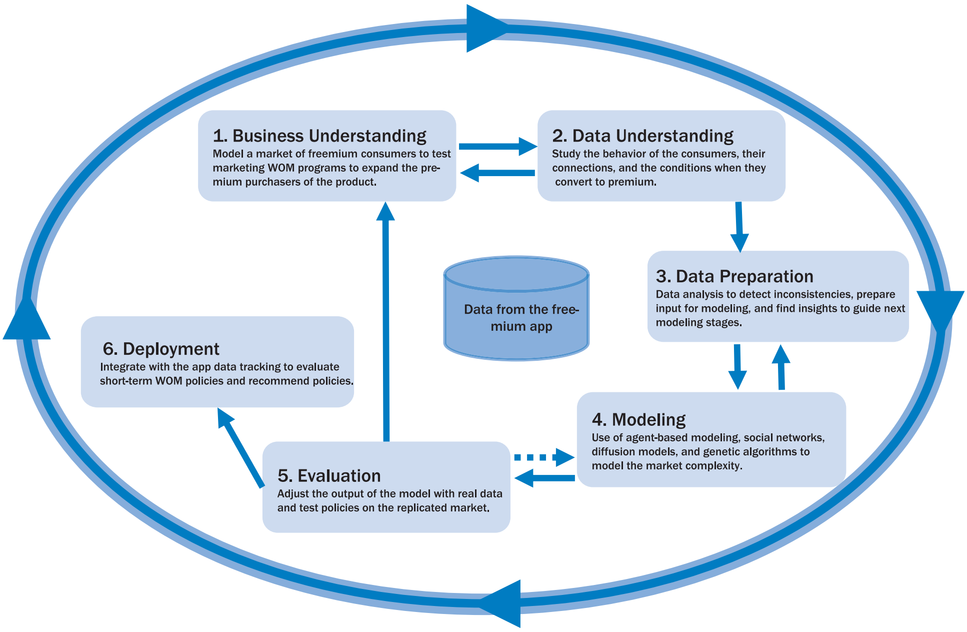
A DSS for WOM Programs

WOM is a powerful force for marketing[50] but managers and marketers in an organization need to know how to design and implement their marketing policies using a WOM program to achieve their business target objectives. Some examples of WOM programs and decisions include the following: balancing WOM with traditional marketing investment[51]; designing influencer strategies on Instagram or Twitter[52]; and harnessing the positive effects of promotions and incentivization campaigns.[53]

It is difficult to know, in advance, whether a WOM campaign will actually result in increased WOM and viral effects. This is why many companies have started to use DSS[54] to help marketers test out WOM programs before the program launches. These DSS can use simulation to provide feedback to marketers and guide them on how best to deploy their marketing strategy. By using a DSS, marketers can explore and test a wide variety of WOM programs and marketing campaigns, observe their impact in a simulated market, and have more knowledge about the market before starting program implementation.

In this case study, we were approached by a massive online freemium app that was interested in knowing what would be the best way to incentivize customer conversions using a WOM campaign.[55] To help them answer their questions, we created a DSS that used an agent-based model designed to simulate WOM dynamics. We then combined this with an ML algorithm, specifically a genetic algorithm, to fit the model to the actual data. Once the model was constructed, we could then use the validated model to explore a wide

FIGURE 2. Diagram showing the main steps for the WOM DSS example.



Note: Analyze the data, design the model, evaluate, and deploy it in a cyclic process. This is a constant cycle since the deployment of each campaign provides better data for improving the model. Dashed arrow shows an additional link between steps 4 and 5 in this case with respect to the standard CRISP-DM cycle. WOM = word-of-mouth; DSS = decision support systems; CRISP-DM = Cross-Industry Standard Process for Data Mining.

variety of marketing policies to determine the optimal policy. We used the CRISP-DM framework to create an AI-based DSS for this system as illustrated in Figure 2.

Business understanding. As background, the online game's main revenue stream is from premium conversions of users who can buy items and extra functionality online.[56] Basic users can freely access and play the game and interact with other users, but premium users receive additional benefits such as weekly in-game currency allowances, the ability to adopt virtual pets, access to all the avatars, and premium-only adventures.

The managers of the organization wanted to know whether WOM played a significant role in the adoption of premium services by freemium users and, based on this knowledge, whether they could potentially design a reward-based marketing campaign to maximize the spread of positive WOM about premium membership, which might, in turn, increase overall adoption rates and revenue. These marketing campaigns involve WOM incentivization of users by rewarding policies (e.g., bonus features for the app or gifts).

The DSS would need to provide answers about which app users to target: the most likely to convert to premium, just random users, or those who are already premium.

Involving stakeholders of the organization at this stage is important and the modeler should spend enough time to understand the business questions. Otherwise, the next steps of the CRISP-DM process will not meet business objectives and decisions might have to be revisited. This project will be judged as successful if:

- The model replicates, using the real data from the company, the network of consumers of the app and their behaviors (including the diffusion models of premium adoption).
- The overall system is able to evaluate incentivization policies and their effect on premium conversions.

Data understanding. When all the goals are defined, it is important to understand the availability and meaning of the relevant data. The organization (in this case, the company that created the online freemium app) provided information about some WOM programs they are interested in exploring, data about the actual consumers and their behaviors, data about consumer relationships within the game, and the global key performance indicators of their business.

Specifically, the data provided by the company had three different dimensions: first, the app use behavior by the users such as daily logins, time from a freemium subscription to acquiring premium subscription, and activity time log; second, information about the friends (contacts) of the app users and activity between these network connections; and, finally, a historical time series of premium conversions as well as new user's registry. The initial data consisted of a dataset of 1.4 million users, with almost 10 million connections between the users (the game supports the idea of "friending" another user). A user was only allowed to have at maximum 100 connections. The dataset was from June 2010 to 2012, and 6.32% of them were premium or became premium during the time of the study.

Data preparation. In the next stage, we carried out exploratory analysis to understand the data. The output of this stage is used to create a DSS that better represents the market reality. The main steps to prepare the available data are the following:

- First, we calculated the conversion rate from freemium to premium. Those active accounts (i.e., users playing for more than 10 days) had a premium conversion rate of 16%.
- We tracked the weekly use of the app and we distinguished two kinds of days: weekdays and weekends. During weekdays, users are not as active as during the weekends because the app is a game for kids and they have more time during the weekends to play with. Also, we observed seasonality during holidays, but the overall trend was stable.

- The average number of friends of the users (average degree of the social network) is 11.8. Premium users have twice as many friends as free users do. Premium users are also more likely to be friends with other premium users.
- The degree distribution of the social network of users is heavily bimodal. There is a group of users who have very few friends, and another group clustered under the upper limit of friends, with fewer users in between these extremes.
- Finally, as the DSS is used for planning short-term campaigns, we extracted periods of time of 2 to 3 months from the total number of weeks to create and validate the behavior of the DSS in the next phases instead of considering large periods of time (e.g., 1 or 2 years of tracked data).

Modeling. The data preparation of the previous stage is closely related to the modeling stage where mathematical and computational tools are employed to create the DSS. As said, the process is cyclic and there is a feedback between them: modeling requires a specific data preparation while the output of the analyzed data also conditions the modeling techniques to use. The modeling techniques we use in this stage can be grouped as follows:

- An ABM framework[57] that generates artificial agents to be the real users of the app. Each agent has a set of behavioral rules[58] that control their activity with the app or subscription state.
- A social network[59] that provides the environment in which the agents operate. It was important to make sure that the network replicated the degree distribution of the real app. To do this, we employed an algorithm that randomly generates links between the agents of the ABM framework until the artificial social network has a given degree distribution.[60] In this case, the algorithm generated a bimodal distribution, as found through the data analysis on the real pool of users of the app. This bimodal distribution was most likely a product of the friend upper limit that the app imposed.
- We included a diffusion mechanism to define and simulate the adoption of premium contents by the users of the app. Given the importance of the social dimension for premium conversions observed in the data analysis, this diffusion mechanism is integrated, together with the activity rules of the agents, as part of the reasoning of the agents. Concretely, two mechanisms of diffusion were modeled: the agent-based Bass model[61] and a complex contagion.[62]
- The framework is also enriched with an automated calibration using genetic algorithms,[63] which is an AI-based optimization procedure. This calibration is also related to the evaluation of the model (next CRISP-DM stage) because it searches for the best set of values for the parameters of the model to fit the key performance indicator that outputs the model with the real historical data (conversions from freemium to premium). In addition, it also helps with the sensitivity analysis of the model and behavioral tests needed for its validation.

Evaluation. It is necessary to evaluate and show the system goodness to guarantee its business success and acceptance by stakeholders. In our case, we examined whether the simulation model captures the reality of the market. It is also important that the system can generate realistic outcomes of a WOM program as the goal is to identify campaigns that minimize the cost of the campaigns and maximize their revenue by means of new premium adopters. A difficulty found when evaluating the deployed system was how to compare the obtained results after applying the policies with respect to the application of other strategies. This is clearly a common problem and therefore we need to rely on simulations' results and past real system behaviors to assess the success of the applied marketing policies.

By using the model built in the previous stage, we evaluate how (by calibrating the parameters with the automated calibration) the output of the model fits with the reality. Deviation measures such as the mean absolute percentage error (MAPE) or the root mean square error (RMSE) are used to quantify the distance of the model with respect to real premium conversions. We follow a train-test approach when calibrating the system: the automated calibration uses around 80% of the period data and leaves 20% for testing the generalization of the model for the data that were not used during the calibration. For instance, to calibrate the model, 60 days of historical premium daily conversions was used as a training set and then 31 days was used for the test dataset.

However, automated calibration is not enough to evaluate the behavior of the system, and techniques such as sensitivity analysis and validation tests are also carried out to study the output of the system. The modeler needs to use automated calibration methods judiciously and in iterative and controlled way in order to manually filter the different alternatives.[64] Otherwise, if modelers blindly accept the calibrated parameters without studying them, these values will be forced to match the historical behavior, with the subsequent risk of treating the model as a black box.

Automated calibration is only a step within model validation and should be considered as part of the model building and validation process. Other useful steps to consider to ensure empirical validation are stress tests and sensitivity analysis (i.e., quantifying how "sensitive" the model is with respect to its input parameters). The latter techniques help us explore parameters that are not working properly or missing features of the modeling. We also employed case studies of incentivization campaigns for users that were premium in order to see how they spread the positive WOM and the implications for increasing the pool of premium users in the artificial market. Sometimes, it became clear that the model needed to be revised and modified (as shown in Figure 2 by the additional dashed arrow between evaluation and modeling). If this was the case, then a previous stage of the CRISP-DM process should be revisited to change the model and evaluate it again.

In addition, the evaluation and first use of the system may also create new questions for the users and modelers about their marketing campaigns (i.e., the business understanding step). In this case, again the ongoing cycle of the CRISP-DM approach will lead to the modification of the modeling, analysis of new data, or creation of new models or sub-models for the DSS. Thus, it is possible to view the CRISP-DM process not as a one-shot process but rather an ongoing cycle that will consider new needs and objectives that are generated every time the process is carried out.

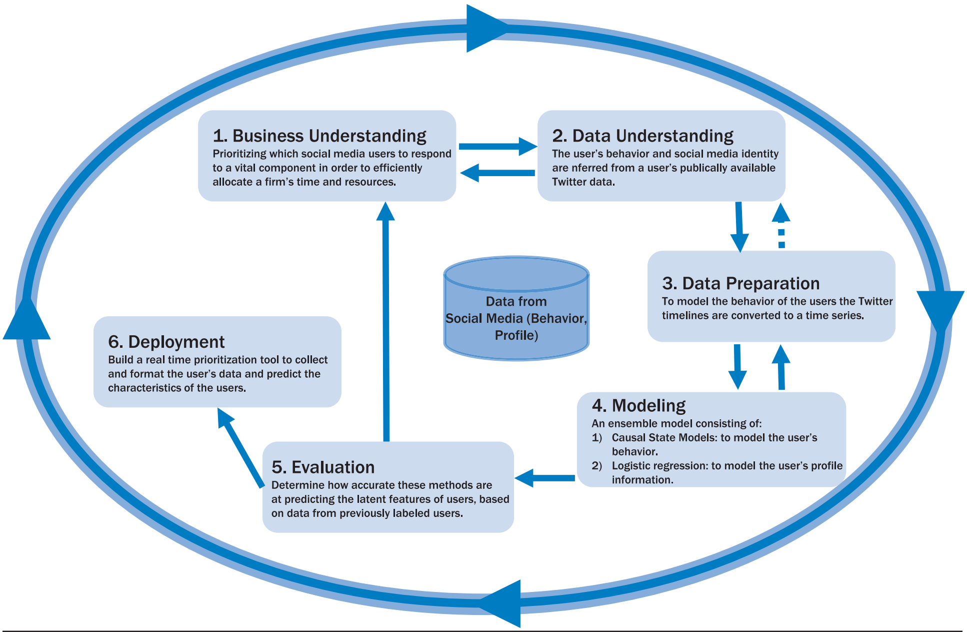
Deployment. Once the DSS is validated, it could be integrated with the marketing department and a periodic tracking of the obtained premium conversions, new users, and their activity is incorporated to the DSS to continuously update and calibrate the system. Managers can then ask questions about the WOM programs and apply them to the system, collecting their outputs and comparing them with the results in the real market (i.e., an in-market validation). These questions were related to the number of targeted users by incentivization policies and how to select them (the most likely to convert to premium, random users, etc.). Having a DSS with on-the-fly recommendations for managing WOM decisions was a real achievement for the managers of the app because they can anticipate and test their marketing ideas with minimal risk.

A manager's questions can be explored by changing or incorporating new initial conditions to the model (e.g., increasing the incentivization of the influences to spread positive WOM), running the models included in the DSS, and comparing with the baseline strategy (not carrying out the campaign or using the standard one). One important finding during this stage was that increasing social influence between users by rewarding users who adopt premium content has a positive nonlinear impact on increasing the number of premium adopters of the app (i.e., referrals can be quite successful). We also observed how the lift in additional premium members does not demonstrate linear behavior when the social influence is increased by rewarding users at the time of adoption. This observation facilitated the managers' understanding of the dynamics of the premium conversions and will guide future directions for testing and applying their app marketing policies.

Automatic Scoring Images for Digital Marketing

In this project, we used visual analytics and AI to understand the role of images in the decision-making process of consumers booking hotels online. The hotel images are important tools to achieve marketing purposes such as creating brand awareness on social media platforms or to facilitate sales. Currently, in many companies, the image selection is done on the basis of "expertise" or a "gut feeling" decision. Brand managers often determine based on their creativity and experience what image to select. Recently, a large global online travel agency was interested in automating this process with AI. So, we applied a combination of multiple CNNs and a support vector regression to score hotel images based on their potential to be clicked on and our algorithm automatically selects the image with the highest potential. In addition, we used the information from our regression model to understand the role of the image in the consumer decision-making

FIGURE 3. Diagram showing the main steps for the image scoring example.



Note: Analyze the data, design the model, and evaluate and deploy it in a cyclic process. Data understanding, data preparation, and modeling happen twice in the process, first for extraction of image information and then image selection. This is illustrated by the additional dashed arrows.

process. The main goal for this project was improving the CTR, but it turns out it is not limited to an increase in CTR. First, by automating the process, we saved marketing managers' time. Second, as a consequence of modeling the impact of images on consumer decisions, we now understand what aspects of an image drive engagement online. Third, image scoring and the selection algorithm can also be offered as a service to hotels. This example will explore the use of CRISP-DM for this Marketing AI project and is illustrated in Figure 3.

Business understanding. The thumbnail, or "champion image," is the first image of a hotel that a consumer encounters, which makes it an important piece of the consumer decision-making process. The "champion image" is currently still randomly chosen and the online travel agency is interested in automating its selection. The technologies from computer science make it easier to extract information from images at a large scale. The information extracted can then be related to the images' success measures. The agency wanted to use these technologies to create an image scoring tool that scores images based on their predicted success in engaging consumers. In this section, we will illustrate the process of creating such a tool using the CRISP-DM framework.

As part of this project, the goal was to develop an image scoring tool to increase the CTR for hotels on the website of the online travel agency. We aimed to create a framework that can be used to automatically extract information from images at a large scale and relate the information extracted to the CTR of the hotel listings. This would then be used to create an image score to automatically select the image with the highest chance of success.

The project will be judged as successful if:

- We can determine the aspects of an image that make it successful in a way that is reusable for the creation and promotion of future images.
- The CTR related to the images increases after selecting images based on the image score.

As mentioned, the data are provided by a global online travel agency, so there is plenty of click-stream data and hotel data available. The most important aspect is to have access to the images and to be able to relate them to the consumers' decision-making process for hotel bookings. Since, as is often the case, these datasets were stored in different locations, we had to work with the company's database specialists to obtain the data from different sources.

Data understanding. Images are very different from traditional relational data. They are a rich source of information, which cannot easily be broken up and stored in a table. It requires special modeling techniques and algorithms to turn images into variables used for analysis. Images in this example consist of 480×480 pixels described by an R, G, and B channel with numbers ranging between 0 and 255, that is, $480 \times 480 \times 3 = 691{,}200$ numbers representing a single image. After processing, we are able to map these data to a series of features that describe the image.

The CRISP-DM framework requires exploration of the data to create an understanding. Images require complicated modeling techniques to extract useful information, which makes data exploration before preparation difficult. We use pre-trained CNNs to extract a rich feature set from the images. Essentially, we use these CNNs to make a prediction about what is portrayed in the image and use that prediction as input for our models. In this Marketing AI setting, the Data Understanding and Data Preparation are a combined process, and even in the early stages, different modeling techniques are evaluated to make the most accurate prediction about the images. First, we follow data understanding, data preparation, and modeling steps to understand the images and then we follow the same three steps again for the image selection process. This process is illustrated in Figure 3 with the additional dashed arrows between steps 2, 3, and 4. These dashed arrows reflect necessary deviations from the standard CRISP-DM framework.

We gathered the "champion" images for all hotels. The "champion" image is the thumbnail that is presented next to the hotel information on the search result page after searching for a destination. We also gathered other hotel information, including aggregate clicks from search result page to hotel page, price of the hotel, number of stars, average customer review, distance to downtown, and related fixed effects. In addition, we gathered customer-level click-stream data that consist of individual searches of consumers on the website. A customer can decide to click on a hotel listing or use one of the actions to sort or filter the search result to set preferences. Naturally, we follow the data exploration step here as well to try to gain an understanding of the behavior of consumers on the webpage.

Data preparation. We extract two types of information from the images: visual complexity and semantic information. Visual complexity captures the complexity of an image, which reflects the overall variation of several aspects of an image. Semantic information covers what is depicted in an image.

The visual complexity of an image is measured by examining the variation at a pixel level. For example, we look at how much variation there is between every pixel in the image for the colors and brightness. The amount of detail in an image is captured by finding the edge density and by the visual clutter of an image.[65] The last element of visual complexity that we consider is the number of objects. We use a CNN to detect the objects and then we simply count how many there are in the image.

For the semantic information of images, we classify images by using two CNNs. The first one[66] returns a distributional representation of 1,000 common objects detected in the image, such as cars, people, and animals, while the second one[67] identifies scene categories, such as beach, hotel room, and library. Once this is done, we have all the information we need from the images ready for analysis and it is time to connect the information from the images with the related metadata. We want to understand the incoming online travelers and understand what information they come across but most importantly what images they see and what makes them click on hotel listings. We connect the weblogs, hotel information, and the images to construct the final dataset.

Modeling. Data preparation and modeling go hand in hand, as some essential steps in understanding the success of images require extracting the information from the image. The next step is to understand the aspect of images that make them successful and to use this information to select the "best" image. There are three stages in the modeling process. First is to model the historic data that we have already gathered. We use these historic data to understand what drives the clicks after the search result for hotel bookings. Second, we use the historic data to train our model to recognize the image with the highest potential for success. Third, we use the trained model to score new incoming images and predict their success.

A regression model allows us to understand the relationship between the CTR and the images. We control for other factors by adding the other variables in our dataset that also impact CTR for online travel search. The second step requires us to create image scores. We can use the results from the regression model and/

or a separate ML model to assign weights to different aspect of images to determine their importance. The ML algorithm will then determine the optimal images to show for an incoming search to maximize CTR.

Evaluation. To test the effect of image score on a particular hotel and as a general model, we need to carry out an experiment. This can be done in an experimental setting where for a certain location or a certain type of booking, we randomly split incoming searches into two possibilities using an A/B test: (A) control group with no changes and (B) treatment group where we show the optimal image based on the image score and keep the rest the same. We are then able to determine whether there is an increase in the CTR for this particular hotel. An increase in CTR means the image scoring method works.

This study highlights some of the aspects of images that make them successful. For instance, are more colors better? Depending on the destination, do consumers like images of beaches or pools better? This information can be used for image design purposes and can help create more effective images. As the images that are used get better and data are collected about CTR on those images, then those data can be used to improve the model even more, creating a model which improves itself over time. We constantly make predictions about what images and what aspects of images work best for an incoming consumer, and with the immediate feedback (clicked or not clicked), we could continuously update the model.

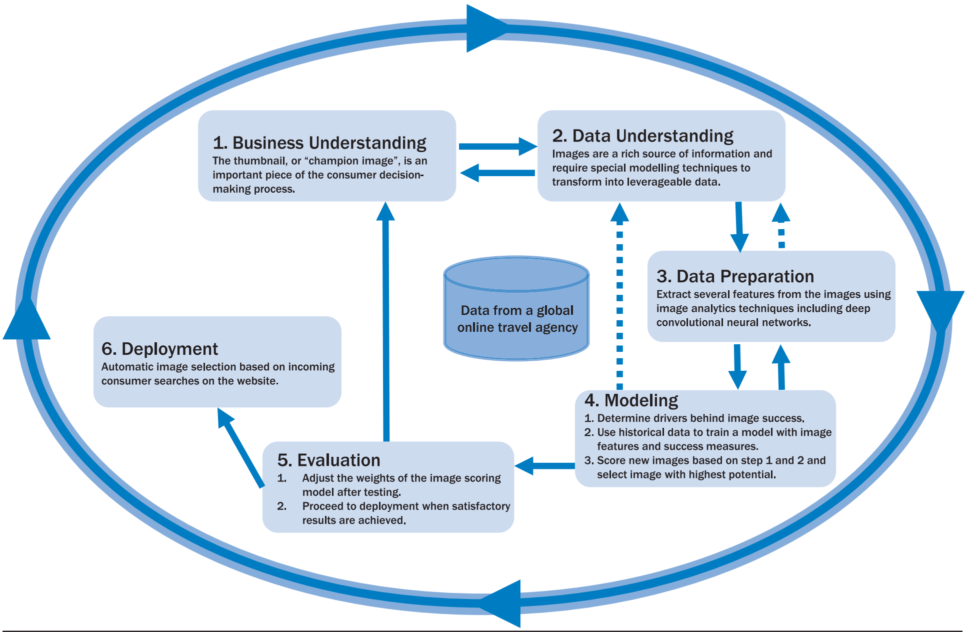
Deployment. A successful deployment of the data mining results requires integration of the model onto the website of the global travel agency and this requires the right flow of data. First, it requires the web developers to integrate the image scoring model into the website, such that for every consumer coming in the right images for the right hotels are shown. It is also essential to have continuous improvement of the model and the weights that determine the importance of the aspects of images to ensure the image selection stays optimal and up-to-date. The deployment should lead to a significant improvement in CTR.

A successful implementation of this automatic image selection would also mean that after going through the different steps of the CRISP-DM framework, the system itself becomes a self-sustainable mechanism. In other words, a true AI image selection tool should be able to go through the data, modeling, and evaluation steps with little human guidance. It would automatically optimize the system and it should make for easy human judgment to complement all the predictions. That is, in many ways, the true goal of Marketing AI.

Prioritizing Customer Service on Social Media

Social media is an important marketing channel for companies to directly engage with consumers and is serving an increasingly important role in customer service.[68] Rather than merely listening, many firms directly engage with the

FIGURE 4. Diagram showing the main steps for the social media customer service example.



Note: Analyze the data, design the model, and evaluate and deploy it in a cyclic process. Data understanding, data preparation, modeling, and evaluation happen twice in the process, first for collecting and formatting the Twitter data, then applying the model to characterize and prioritize the users.

consumers through social media platforms. This has led to an increase in direct interactions between the firms and individual consumers that has revolutionized customer relationship management (CRM), creating a type of social CRM.[69] Firms are deploying resources to help respond to these concerns, but most firms do not have the resources to respond to every customer service comment that comes in over social media, so it is important to prioritize these issues. To effectively determine which users to respond to on social media, we need to determine which features are important and whether we can accurately detect these characteristics.

Ma et al. showed that it was useful to respond to customers on social media, but that responding to them may encourage future negative WOM.[70] Thus, being able to identify which users should be responded to and allocate the resources to engage with these consumers is very important. This requires building a tool that can identify these users in near real time and provide a more efficient allocation of time and resources. The general goal of this project was to develop a Marketing AI tool to help prioritize customer service requests on social media and is illustrated in Figure 4. This marketing AI tool consists of analyzing and summarizing the social media users into a Markovian-like model and building a supervised learning model to classify and prioritize the users. Unlike the previous examples, this project was undertaken without a focal firm, but instead as a general tool development project.

Business understanding. To break down the overall perspective of prioritizing social media users, it was first necessary to consider what aspects of a user would mean that they are a higher priority. After examining the business objectives at hand, it appeared that there were three features of the users that could be ascertained from social media that could help determine their value to the firm:

- Are they geographically relevant customers, for example, are they in a market where the firm's product/service is sold?
- Do they have a potentially high CLV?
- Will they provide an overall increase in positive WOM about the brand?

The first characteristic reflects whether the user can even be a customer of the firm. If a user cannot be a customer due to geographical constraints, then the only potential value is related more to brand building than actual sales (geography). The second characteristic focuses on the user's potential to be a high CLV consumer (CLV). A wealthy user has the possibility of being a valuable customer. The third feature is an estimate of the user's effect on the firm's digital WOM. It has been well documented that the digital WOM affects the firm's marketing effects and sales[71] and is an important feature to include when prioritizing users.

The project will be judged as successful if:

- We can accurately classify social media users into their appropriate groups based on the user's profile information and behavior.
- We can use that classification to prioritize social media customer service requests in a manner that optimizes a company's resources.

Data understanding. This Marketing AI application focuses on Twitter. Twitter was chosen because it is one of the most active platforms where consumers voice their opinions about brands.

Firms are able to directly respond to users' complaints and concerns on Twitter, and many users take advantage of this capability.[72] For each one of these users, there are many variables to consider that can be extracted from social media. The approach presented in this example uses only publicly available social media content. Using only publicly accessible data ensures that any firm can collect the data and implement these methods.

For the initial development of the project, we decided to focus on the two types of data that seemed to have the best potential for helping with classification: profile information and behavioral data. Profile data are data available on the profile of Twitter public accounts and include information such as the number of tweets, the number of followers, and the number of Twitter accounts who follow the user. The behavioral data for a user is their activity on Twitter over time. The benefit of using these data as the independent variables to determining the classification of a user is that these variables can be collected and formatted for analysis at close to real time, which was important for the business context explained above. The ability to collect these data at near real time allows for the customer service to quickly collect the data, apply the model, and prioritize users.

Data preparation. To prepare the data, both the profile information and the Twitter timeline of the users need to be formatted. Using the Twitter API, these datasets can be collected and formatted relatively quickly, minimizing the time between the initial customer service interaction and the firm's response.

The profile information includes six variables of interest. These variables are as follows: *number of tweets* by the user; *number of followers; number of followees*, that is, the number of Twitter accounts that follow the user; *number of statuses favorited* by the user; whether the user is *verified*, that is, a user that Twitter has confirmed that the user is who they claim to be; and *the number of times the user is listed*, that is, appears in a public Twitter list. These profile variables are not only publicly accessible, but they are also required for all users. Thus, all of the users being investigated will contain these variables. These variables can be used as the independent variables to build models to classify the users into the latent groups: CLV, Location, and WOM.

The API gives companies the ability to quickly collect the timeline of the last 3,200 tweets from a user. Using these tweets, we format the data into time series, consisting of whether or not a user tweeted during a two-hour window. Once all of these data are formatted, the next step is to develop the model.

Modeling. Once the data are formatted, we can apply the AI methods to classify the users based on the features of interest. In the end, the decision was made to use two separate learning algorithms for the two types of data, profile and behavior data, and then combine the results into an ensemble model to classify the user. Combining the two types of data improved the performance of the methods compared with a single model.

For the *profile* information, it was decided to use a traditional logistic model since there was a discrete class output and static independent variables. For the *behavior* data, a relatively new technique known as Causal State Models (CSMs) was used. CSMs have been selected to model the time series as opposed to other time-series models such as ARIMA (Autoregressive Integrated Moving Average) models, because CSMs provide a clear easy interpretable model. The resulting CSM provides an exact model where the researcher can determine the current state of the user based on the previous events. The CSMs for each user are constructed through the Causal State Splitting and Reconstruction algorithm,[73] which produces a CSM that is minimally complex while maintaining the maximum predictive power.[74] Once the CSMs are constructed for a user, we apply K-Nearest Neighbor method to classify the user into the appropriate level. Using a distance metric based on the differences between the probabilities of events occurring between the two models, we are able to find the K-nearest CSMs from the training set of users. The user is then classified by the majority of these K-nearest neighbors.

The final classification of the user, which accounts for both of these types of model, is a weighted average of the results of these two models. This weighted average improves classification accuracy of the users while accounting for the two different types of data.

Evaluation. To evaluate the proposed methods, 4,776 users, who reached out to a company on a specific day, were investigated. For each one of these users, their profile information and Twitter timelines were collected. These users were selected because each one of these users directly tweeted at a firm using the official Twitter handle of that firm (called *mentions* and *replies* on Twitter), on a specific date. The collection of these user's timelines and profile information were collected on the same day in order to ensure the same research time frame for each user. These data were collected seven months after the initial tweets to the firms. This seven-month delay was necessary so we could determine what the activity of the user would be after their initial contact with the firm. To assess the performance of the models in all cases, the dataset was divided into a training set and a testing set, with $5/6$ of the users in training set and the remaining $1/6$ users being in the testing set. The models were built on the training set, and the testing set was used to determine the performance of the models.

Each of these users was labeled into their appropriate classes for each of the three characteristics of interests. As stated earlier, the three prioritizing latent characteristics investigated by this study are the location of the user, the customer's lifetime value, and the brand WOM of the user. The success of the classification method is dependent on which feature was being predicted. The model had the most success in predicting the location of the user and the Internet brand WOM of the user. The model struggled with predicting the customer's lifetime value. Despite the lack of success in predicting the customer's lifetime value, being able to detect the user's brand WOM and location of the user provides the firm with useful information when deciding which users to respond to on social media.

Deployment. A successful deployment of this system would involve integrating the model into customer service systems. The brand's Twitter account would have to collect the user's content whenever multiple users reach out to the firm on Twitter. The model will then help classify the user according to the latent characteristics described above. Based on these latent classes and company information, a final tool could be developed that would present customer service representatives with a list of users assigned to them for response prioritized by this information. Each firm might prioritize different aspects of the classes, for example, one firm might want to prioritize high CLV over WOM and geography. Similar methods could also be developed for other classes of interest besides CLV, Geography, and WOM, if the firm was interested in those classes. The exact same models would not be used, but new models could easily be generated using the same basic framework.

Conclusion

We have shown how the CRISP-DM framework can be utilized to develop AI solutions to marketing problems. We have illustrated this idea with three novel and interesting case studies that are significant advances in Marketing AI themselves and have elucidated a number of principles and concerns that marketing managers should address when carrying out Marketing AI projects.

There are a few lessons that we would like to highlight for future efforts in Marketing AI. We will break them down by the relevant phase:

- *Business understanding*: Involving stakeholders of the organization is important and the modeler should spend enough time to understand the business questions to answer. There should be a clear understanding of what the business problem is, why it is an important problem to solve, and what a good solution will look like. Otherwise, the next steps of the CRISP-DM process will not be adequate, and previous decisions and steps might need to be revisited again.
- *Data understanding*: Our understanding of data has become more complex as we have developed ways to deal with new forms of data, such as images, text, and video. Often these data forms require going through some of the CRISP-DM steps twice. First, we follow the data understanding, data preparation, modeling, and evaluation steps to make predictions about the unstructured data and turn this into features and representations that we can use in our models, and then we follow the same steps again using these new features to actually build our Marketing AI solution.
- *Data preparation*: One of the most important aspects of data preparation is constructing and enhancing the data. Many modeling forms work better when the raw data have been transformed in some way. For instance, CSMs require discretized data, and the logistic models of the image selection example needed some aspects of the images to be transformed into arrays of Boolean variables to work properly. If this process is carried out correctly, it is possible to transform data that were not usable into something that can help increase the success of marketing outcomes.
- *Modeling*: The Marketing AI process is cyclic and there is a feedback between the phases. Modeling requires a specific data preparation while the output of the analyzed data also determines which modeling technique is best to use. In addition, modeling is often required to make predictions about complex data formats, which in turn is used as input again into another model.
- *Evaluation*: Calibration is only a step within model validation and should be considered as part of the model building and validation process. Other useful steps to consider to ensure empirical validation are stress tests, case studies, and sensitivity analysis.
- *Deployment*: The whole CRISP-DM process might be thought of as a never-ending cycle. Each iteration creates new questions and new possibilities, which can be addressed in the next cycle. The goal of Marketing AI is to automate as much as possible the continual refinement of the previous models, so

that even as the models themselves answer more and more questions, they are also answering them better and better over time.

Authors' Note

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Laboratory for Analytic Sciences (LAS) and/or any agency or entity of the U.S. Government.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS).

Author Biographies

Gijs Overgoor is a Marketing PhD candidate in the Amsterdam Business School at the University of Amsterdam and a visiting scholar at the Poole College of Management at North Carolina State University (g.overgoor@uva.nl).

Manuel Chica is Ramon y Cajal Senior Researcher at the University of Granada, Spain, and conjoint lecturer at the University of Newcastle, Australia (manuelchica@ugr.es).

William Rand is an associate professor of Marketing in the Poole College of Management at North Carolina State University (wmrand@ncsu.edu).

Anthony Weishampel is a Statistics PhD candidate in the Department of Statistics at North Carolina State University (acweisha@ncsu.edu).

Notes

1. Adnan Darwiche, "Human-Level Intelligence or Human-Level Intelligence or Animal-Like Abilities?" *Communications of the ACM*, 61/10 (October 2018): 56-67.
2. Hal Conick, "The Past, Present and Future of AI in Marketing," *Marketing News*, 51/1 (January 2017): 26-35.
3. Oded Netzer, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko, "Mine Your Own Business: Market-Structure Surveillance through Text Mining," *Marketing Science*, 31/3 (May/June 2012): 521-543; Seshadri Tirunillai and Gerard J. Tellis, "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research*, 51/4 (August 2014): 463-479.
4. Geng Cui, Man Leung Wong, and Hon-Kwong Lui, "Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming," *Management Science*, 52/4 (April 2006): 597-612.
5. Aurélie Lemmens and Christophe Croux, "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, 43/2 (May 2006): 276-286.
6. John R. Hauser, Guilherme Liberali, and Glen L Urban, "Website Morphing 2.0: Switching Costs, Partial Exposure, Random Exit, and When to Morph," *Management Science*, 60/6 (June 2014): 1594-1616; John R. Hauser, Glen L. Urban, Guilherme Liberali, and Michael Braun, "Website Morphing," *Marketing Science*, 28/2 (March/April 2009): 202-223.

7. Pradeep Chintagunta, Dominique M. Hanssens, and John R. Hauser, "Marketing Science and Big Data," *Marketing Science*, 35/3 (May/June 2016): 341-342.
8. Colin Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing*, 5/4 (Fall 2000): 13-22.
9. Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Boston, MA: Pearson, 2018).
10. Ibid.
11. Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Boston, MA: Harvard Business Review Press, 2018).
12. Tom M. Mitchell, *Machine Learning* (New York, NY: McGraw-Hill, 1997).
13. Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective* (Cambridge, MA: MIT Press, 2012).
14. Agrawal et al., op. cit.
15. Marno Verbeek, *A Guide to Modern Econometrics* (Chichester: John Wiley, 2008).
16. Hal R. Varian, "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28/2 (Spring 2014): 3-28.
17. Galit Shmueli, Peter C. Bruce, Nitin R. Patel, Inbal Yahav, and Kenneth C. Lichtendahl Jr., *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R* (Hoboken, NJ: John Wiley, 2017).
18. Alina Selyukh, "Optimized Prime: How AI and Anticipation Power Amazon's 1-Hour Deliveries," *NPR*, November 21, 2018, https://www.npr.org/2018/11/21/660168325/optimized-prime-how-ai-and-anticipation-power-amazons-1-hour-deliveries.
19. Shmueli et al., op. cit.
20. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (Cambridge, MA: MIT Press, 2016).
21. Nikolay Burlutskiy, Miltos Petridis, Andrew Fish, Alexey Chernov, and Nour Ali, "An Investigation on Online versus Batch Learning in Predicting User Behaviour," in *Research and Development in Intelligent Systems XXXIII. Conference Paper at International Conference on Innovative Techniques and Applications of Artificial Intelligence*, ed. M. Bramer and M. Petridis (New York, NY: Springer, 2016), pp. 135-149.
22. Verbeek, op. cit.
23. Nello Cristianini and John Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge, UK: Cambridge University Press, 2000).
24. William Rand and Roland T. Rust, "Agent-Based Modeling in Marketing: Guidelines for Rigor," *International Journal of Research in Marketing*, 28/3 (September 2011): 181-193; Uri Wilensky and William Rand, *Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo* (Cambridge, MA: MIT Press, 2015).
25. Ajay K. Agrawal, Joshua S. Gans, and Avi Goldfarb, "Prediction, Judgment and Complexity: A Theory of Decision Making and Artificial Intelligence," NBER Working Paper No. 24243, National Bureau of Economic Research, Cambridge, MA, 2018.
26. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay, "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, 12 (2011): 2825-2830.
27. Bernd Bischl, Michel Lang, Lars Kotthoff, Julia Schiffner, Jakob Richter, Erich Studerus, Giuseppe Casalicchio, and Zachary M. Jones, "mlr: Machine Learning in R," *Journal of Machine Learning Research*, 17/170 (2016): 1-5.
28. Terry Therneau, Beth Atkinson, and Brian Ripley, "rpart: Recursive Partitioning and Regression Trees," r Package Version 4.1-10, 2015.
29. Hadley Wickham, Romain Francois, L. Henry, and K. Müller, "dplyr: A Grammar of Data Manipulation," r Package Version 0.5. 0, 2016.
30. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," Software, 2015, tensorflow.org.

31. François Chollet et al., Keras, https://github.com/fchollet/keras, 2015.
32. Eric Jones, Travis Oliphant, Pearu Peterson, et al., "SciPy: Open Source Scientific Tools for Python," 2001, https://www.scipy.org/citing.html.
33. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, "The WEKA Data Mining Software: An Update," *ACM SIGKDD Explorations Newsletter*, 11/1 (2009): 10-18; Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan, "Orange: Data Mining Toolbox in Python," *Journal of Machine Learning Research*, 14 (August 2013): 2349-2353.
34. J. Alcalá-Fdez, L. Sánchez, S. García, M. Del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. C. Fernández, and F. Herrera, "KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems," *Soft Computing*, 13/3 (2009): 307-318.
35. Shearer, op. cit.
36. Shmueli et al., op. cit.
37. Umair Shafique and Haseeb Qaiser, "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)," *International Journal of Innovation and Scientific Research*, 12/1 (November 2014): 217-222; Meta S. Brown, "What IT Needs to Know about the Data Mining Process," *Forbes*, July 29, 2015; Godfrey Onwubolu, "An Inductive Data Mining System Framework," in *Proceedings of the International Workshop on Inductive Modeling (IWIM '09)* (Krynica, Poland: Citeseer, 2009), pp. 108-113.
38. Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth, "CRISP-DM 1.0 Step-by-Step Data Mining Guide," Technical Report, SPSS, 2000, https://www.scribd.com/document/264461662/CRISP-DM-1-0-Step-By-Step-Data-Mining-Guide.
39. J. A. Harding, M. Shahbaz, Srinivas, and A. Kusiak, "Data Mining in Manufacturing: A Review," *Journal of Manufacturing Science and Engineering*, 128/4 (2006): 969-976.
40. B. Carneiro da Rocha and R. Timóteo de Sousa Júnior, "Identifying Bank Frauds Using CRISP-DM and Decision Trees," *International Journal of Computer Science and Information Technology*, 2/5 (October 2010): 162-169.
41. Sérgio Moro, Raul Laureano, and Paulo Cortez, "Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology," in *Proceedings of European Simulation and Modeling Conference—ESM 2011*, EUROSISETI, 2011, pp. 117-121, https://www.semanticscholar.org/paper/Using-data-mining-for-bank-direct-marketing%3A-an-of-Moro-Laureano/1999417377ec21ecf7f7f55af62975065f785fb2; Wendy Gersten, Rüdiger Wirth, and Dirk Arndt, "Predictive Modeling in Automotive Direct Marketing: Tools, Experiences and Open Issues," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM, 2000), pp. 398-406, https://dl.acm.org/citation.cfm?id=347174.
42. Michael S. Garver, "Using Data Mining for Customer Satisfaction Research," *Marketing Research*, 14/1 (2002): 8.
43. Alon Halevy, Peter Norvig, and Fernando Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems*, 24/2 (2009): 8-12.
44. Ronen Feldman and James Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* (Cambridge, UK: Cambridge University Press, 2007).
45. Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques* (Cambridge, MA: Morgan Kaufmann, 2016).
46. Murphy, op. cit.
47. Jerome Friedman, Trevor Hastie, and Robert Tibshirani, *The Elements of Statistical Learning* (New York, NY: Springer-Verlag, 2001).
48. Claudia Beleites, Ute Neugebauer, Thomas Bocklitz, Christoph Krafft, and Jürgen Popp, "Sample Size Planning for Classification Models," *Analytica Chimica Acta*, 760 (2013): 25-33.
49. Friedman et al., op. cit.
50. Michael Trusov, Randolph E. Bucklin, and Koen Pauwels, "Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73/5 (September 2009): 90-102.
51. Barak Libai, Eitan Muller, and Renana Peres, "Decomposing the Value of Word-of-Mouth Seeding Programs: Acceleration versus Expansion," *Journal of Marketing Research*, 50/2 (April 2013): 161-176.
52. Iman Golkar Amnieh and Marjan Kaedi, "Using Estimated Personality of Social Network Members for Finding Influential Nodes in Viral Marketing," *Cybernetics and Systems*, 46/5 (2015): 355-378.

53. Philipp Schmitt, Bernd Skiera, and Christophe Van den Bulte, "Referral Programs and Customer Value," *Journal of Marketing*, 75/1 (2011): 46-59.
54. Gary L. Lilien, "Bridging the Academic-Practitioner Divide in Marketing Decision Models," *Journal of Marketing*, 75/4 (2011): 196-210.
55. Manuel Chica and William Rand, "Building Agent-Based Decision Support Systems for Word-of-Mouth Programs. A Freemium Application," *Journal of Marketing Research*, 54/5 (October 2017): 752-767.
56. Ibid.
57. Charles M. Macal and Michael J. North, "Tutorial on Agent-Based Modeling and Simulation," *Proceedings of the 37th Conference on Winter Simulation* (New York, NY: ACM, 2005), pp. 2-15, https://dl.acm.org/citation.cfm?id=1162712.
58. Wilensky and Rand, op. cit.
59. Mark Newman, Albert-László Barabási, and Duncan J. Watts, *The Structure and Dynamics of Networks* (Princeton, NJ: Princeton University Press, 2006).
60. Fabien Viger and Matthieu Latapy, "Efficient and Simple Generation of Random Simple Connected Graphs with Prescribed Degree Sequence," *Lecture Notes in Computer Science, Computing and Combinatorics*, 3595 (2005): 440-449.
61. Rand and Rust, op. cit.
62. Damon Centola and Michael Macy, "Complex Contagions and the Weakness of Long Ties," *American Journal of Sociology*, 113/3 (November 2007): 702-734.
63. Manuel Chica, José Barranquero, Tomasz Kajdanowicz, Sergio Damas, and Óscar Cordón, "Multimodal Optimization: An Effective Framework for Model Calibration," *Information Sciences*, 375 (2017): 79-97.
64. Ibid.
65. Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano, "Measuring Visual Clutter," *Journal of Vision*, 7/2 (August 2007): 1-17.
66. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (New York, NY: IEEE, 2016), pp. 770-778, https://ieeexplore.ieee.org/document/7780459.
67. Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba, "Places: A 10 Million Image Database for Scene Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40/6 (June 2018): 1452-1464.
68. Andrew Rohm, Velitchka Kaltcheva, and George Milne, "A Mixed-Method Approach to Examining Brand-Consumer Interactions Driven by Social Media," *Journal of Research in Interactive Marketing*, 7/4 (2013): 295-311; Neil Woodcock, Nick Broomfield, Geoff Downer, and Michael Starkey, "The Evolving Data Architecture of Social Customer Relationship Management," *Journal of Direct, Data and Digital Marketing Practice*, 12/3 (January 2011): 249-266; Liye Ma, Baohong Sun, and Sunder Kekre, "The Squeaky Wheel Gets The Grease—An Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science*, 34/5 (2015): 627-645.
69. Woodcock et al., op. cit.
70. Ma et al., op. cit.
71. Judith Chevalier and Dina Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marking Research*, 43/3 (August 2006): 345-354; Wenjing Duan, Bin Gu, and Andrew Whinston, "Do Online Reviews Matter?—An Empirical Investigation of Panel Data," *Decision Support Systems*, 45/4 (November 2008): 1007-1016; Ma et al., op. cit.
72. Sabine Einwiller and Sarah Steilen, "Handling Complaints on Social Network Sites—An Analysis of Complaints and Complaint Responses on Facebook and Twitter Pages of Large US Companies," *Public Relations Review*, 41/2 (June 2015): 195-204.
73. Cosma Shalizi and Kristina Shalizi, "Blind Construction of Optimal Nonlinear Recursive Predictors for Discrete Sequences," *CoRR*, cs.LG/0406011, 2004.
74. David Darmon, Jared Sylvester, Michelle Girvan, and William Rand, "Understanding the Predictive Power of Computational Mechanics and Echo State Networks in Social Media," *arXiv: 1306.6111*, 2013.