# Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics

Jose Ramon Saura

*Department of Business Economics, Rey Juan Carlos University, Madrid, Spain*

## ARTICLE INFO

## ABSTRACT

In the last decade, the use of Data Sciences, which facilitate decision-making and extraction of actionable insights and knowledge from large datasets in the digital marketing environment, has remarkably increased. However, despite these advances, relevant evidence on the measures to improve the management of Data Sciences in digital marketing remains scarce. To bridge this gap in the literature, the present study aims to review (i) methods of analysis, (ii) uses, and (iii) performance metrics based on Data Sciences as used in digital marketing techniques and strategies. To this end, a comprehensive literature review of major scientific contributions made so far in this research area is undertaken. The results present a holistic overview of the main applications of Data Sciences to digital marketing and generate insights related to the creation of innovative Data Mining and knowledge discovery techniques. Important theoretical implications are discussed, and a list of topics is offered for further research in this field. The review concludes with formulating recommendations on the development of digital marketing strategies for businesses, marketers, and non-technical researchers and with an outline of directions of further research on innovative Data Mining and knowledge discovery applications.

## Introduction

Since the beginning of the 21st century, both Digital Marketing (DM) and Data Sciences (DS) have remarkably evolved in terms of use and profitability (Tiago & Veríssimo, 2014). This has led to the emergence of a digital ecosystem, which connects users 24/7 and which has shaped users' new habits and behaviors (Mayer-Schönberger & Cukier, 2013).

DM is defined as a set of techniques developed on the Internet with to persuade users to buy a product or service (Avery, Steenburgh, Deighton, & Caravella, 2012). Today, the daily roadmap for companies that operate on the Internet includes techniques such as Search Engine Optimization (SEO), i.e. optimization of search results from major search engines; Search Engine Marketing (SEM) or programmatic advertising, i.e. strategies to sponsor ads in search engines or in advertising space on banners in websites; as well as Social Media Marketing (SMM), i.e. strategies of interacting with users on social networks through social ads (Lies, 2019; Palos-Sanchez, Saura, & Martin-Velicia, 2019).

In recent years, DM has spurred a considerable research interest among scholars (Kannan, 2017). For instance, Rogers and Sexton

(2012) sought to understand the key ways to improve profitability or ROI (Return of Investment) in DM. Furthermore, Kumar et al. (2013) measured the influence of data on the DM ecosystem. Likewise, Saura, Palos-Sánchez, and Cerdá Suárez (2017) identified the metrics to measure the efficiency of each of the DM actions developed by a company on the Internet.

Numerous studies demonstrated that a key way to increase the effectiveness of DM strategies is the application of DS techniques in this industry (Braverman, 2015; Dremel, Herterich, Wulf, & Vom Brocke, 2020; Sundsøy, Bjelland, Iqbal, & de Montjoye, 2014). For example, Kelleher and Tierney (2018) argued that DS can increase the effectiveness of DM by improving issues such as (i) companies' management of the information collected from users; (ii) the type and source of data from the companies' datasets, and (iii) application of new data analysis and innovative techniques to create knowledge (Palacios-Marqués et al., 2016).

Furthermore, Fan, Lau, and Zhao (2015) and Saura and Bennett (2019) underscored the importance of several important aspects, such as the type of data collected from different online sources, purchases made by users, and their digital habits or behaviors. Likewise, Wedel and Kannan (2016) demonstrated that, in order to increase the chances of success on digital and social media platforms, companies should identify unsuspected patterns using Artificial Intelligence (AI) or Machine Learning (ML) techniques. Accordingly, the DM industry has been increasingly influenced by

*E-mail address:* joseramon.saura@urjc.es

research areas such as Information Sciences (IS) or Computer Sciences (CS), as well as by all other areas of research that facilitate collecting, ordering, and managing data (Provost & Fawcett, 2013).

Until now, the key tasks of DS have included improving the storage capacity of company data, performing market research and consumer segmentation, or extracting key information regarding company problems (Loebbecke & Picot, 2015). However, DS is a broad ecosystem that encompasses different pattern identification strategies, models of analysis, performance indicators, statistical variables, and technicalities skills linked to a great technological expertise (Leeflang, Verhoef, Dahlström, & Freundt, 2014). However, several studies have highlighted that there is a skills gap in this industry (e.g., Ghotbifar, Marjani, & Ramazani, 2017; Royle & Laing, 2014). Specifically, both marketers without expertise in IS, CS, or DS and non-technical researchers who lack the knowledge of data management have faced the challenge of acquiring such knowledge and skills and using them not only technically, but also strategically and operationally. These challenges and the ways to overcome them are the motivation of the present study.

In order to analyze the impact of the increase in companies' use of DS on DM, the present study performs an in-depth review of (i) methods of analysis, (ii) uses, and (iii) performance metrics based on DS applied in the scientific literature to DM strategies. The aforementioned three aspects will be studied, explained, and analyzed from the marketer's point of view, rather than from that of a data scientist. By reviewing the main concepts of DS framework applied to DM, the present study will allow marketers and non-technical researchers to better understand the main applications of DS to DM, as well as to become more aware of the importance of each such application.

Considering the scarcity of previous research on the relationship between DS with DM, the present review both bridges a gap in the literature and offers directions of further research in this area. The results of this review will allow marketers and non-technical researchers to better understand how the DS ecosystem applied to DM works, and what the key points of and alternatives to applying these techniques to DM are.

Therefore, the present study addressed the following two research questions:

RQ1: What are the main methods of analysis, uses, and performance metrics of Data Sciences applied in Digital Marketing?

RQ2: What are the areas of further research on the use of Data Science in Digital Marketing?

To address these two questions, a Systematic Literature Review (SLR) is performed based on the publications available from several scientific databases, such as ACM Digital Library, AIS Electronic Library, IEEE Explore ScienceDirect, and Web of Sciences.

The remainder of this paper is structured as follows. The following section presents the theoretical framework of DS. In the second part, Theoretical background, the main concepts of the DS framework are presented. Sections "Methodology development" and "Analysis of results" present the methodology and report the results, with a particular focus on the development of descriptors that define the key aspects of DS in DM. Finally, the last section draws conclusions and outlines directions of further research.

## Data science: theoretical framework

Overall, the major goal of DS is to extract knowledge from data analysis to answer specific research questions (Kelleher & Tierney, 2018). By analyzing the data, DS techniques make it possible to extract patterns from databases to explain a problem or to formulate hypotheses. In DS, a key idea is that the patterns identified in the data are (i) non-obvious and (ii) useful for companies (Berry & Linoff, 2004).

**Table 1**
Types of patterns and descriptions in Data Sciences.

| Type of patterns | Patterns description |
|---|---|
| Clustering | Known as segmentation in business, clustering consists of the identification of behaviors, tastes, or habits that identify the same group of consumers among other uses |
| Association rule mining (ARM) | ARM consists of identifying patterns of products purchased at the same time |
| Outlier detection/anomaly | Identification of unusual or uncommon patterns that could be fraudulent, such as in financial or insurance sectors |
| Prediction patterns (PP) | A pattern that predicts a missing value of an attribute (product, service, etc.) instead of predicting the future |

In this relation, it is important to note that, in terms of detecting patterns, humans can identify a maximum of three attributes, or characteristics of an item (product, services, community, etc.). These attributes are also known as features or variables (Saura & Bennett, 2019). However, with DS patterns, hundreds and thousands of attributes (variables) can be simultaneously identified (Berry & Linoff, 2004).

Patterns identified with DS techniques help to obtain actionable insights, i.e. what researchers or data scientists want to extract from the identified patterns (Kelleher, Mac Namee, & D'arcy, 2015). Therefore, the term 'insight' is this context refers to the capacity of patterns to provide meaningful information that can help to solve the problem at stake. The word 'actionable' here means that insights extracted from patterns can in some way be used by the company (Davenport, 2014). In DS, there are different types of patterns that can be applied to the DM industry (see Table 1).

Depending on the company's goals when developing DM, different types of patterns can be used to improve these strategies, as well as to enhance the company's ability to understand and structure the main attributes, features, or variables extracted from companies' databases (Berry & Linoff, 2004). In this sense, the type of collected data is also important, from the company's perspective, as the developed strategies that target digital platforms and social networks should be data-driven (Shareef, Kapoor, Mukerji, Dwivedi, & Dwivedi, 2020).

Large volumes of data are referred to as Big Data (BD) (Gandomi and Haider, 2015). Big Data are characterized by the three Vs: (i) *volume*, i.e. excessive amounts of data, (ii) *variety* of data types, and (iii) *velocity* at which the data must be processed (Kelleher & Tierney, 2018). The foundation of BD as they exist today was laid down by Codd's (1971) Relational Data Model (RDM) that allowed for collecting and storing information, as well as making direct queries on information in databases. This advance removed the concern of the physical location of a database. This was an important milestone in the DS industry—previously, databases were in separate physical storage (Dwork & Roth, 2014).

Codd (1971) also laid the foundations of Structured Query Language (SQL), the current standard for querying databases. The latest developments in data storage have led to the generation of new databases known as NoSQL databases. NoSQL databases store variable data and their attributes with languages such as JavaScript Object Notation (JSON). JSON weighs less, has a higher processing speed, and is self-describing, rather than based on table-based relational model like SQL. Today, data warehouses (DW) are more easily available for analysis, measurement, and control ( Janssen et al., 2020).

The origin and sources of the data are also important in DS. Depending on the origin of the data, different types of DS-based approaches are available for each type of analysis. Table 2 shows

**Table 2**
Types of data sources and descriptions in Data Sciences applied to Digital Marketing.

| Type of data | Data description |
| --- | --- |
| Transactional data | Information regarding sales, invoices, receipts, shipments, payments, insurance, rentals, etc |
| Non-transactional data | Demographic, psychographic, behavioral, lifestyle data, etc |
| Operational data | Data on strategies and actions related to logistics and business operations |
| Online data (Sources) | User Generated Content (UGC), emails, photos, tweets, likes, shares, websites, web searches, videos, online purchases, music, etc |

**Table 3**
Datasets indicators in Data Sciences.

| Dataset | Indicator description |
| --- | --- |
| Variable | Denotes an individual abstraction in the dataset |
| Attribute | Attributes are generally composed of numerical, ordinal, and nominal values. Attributes are characteristics of a variable |
| Feature | Features respond to the characteristics of variables such as attributes |
| Entity | Entity is normally defined with a number of attributes |
| Instance | In DS terms and indicators like instance, the terms example, entity, object, case, individual or record could be used in the DS literature to refer to a row |

**Table 4**
Main machine learning models.

| Type | Description |
| --- | --- |
| Ensemble models (EM) | EM are used to make predictions using a model source where each model votes on each query |
| Deep-learning neural networks | Deep-learning neural networks have different layers of networks that discover patterns and can be trained. Models can also learn these patterns from complex datasets and can be applied to others later by following the learned criteria |
| Machine Vision (MV) | Within the deep-learning neural networks, MV allows for a visual identification and recognition of objects, people, products, etc |
| Natural-Lenguage Processing (NLP) | Based on text analysis and predefined patterns, NLP works with language and text to identify insights that explain patterns non-identifiable by humans |

**Table 5**
Machine learning approaches applied to Digital Marketing.

| Type | Description |
| --- | --- |
| Supervised learning (SL) | SL the action of the ML that allows an algorithm to map an input to an-output normally known as input–output pairs |
| Unsupervised learning (SL) | SL is the action of the ML that allows the function of identify for previously undetected patterns in a dataset with no pre-existing labels |
| Support Vector Machines (SVM) | Support vector is a one-class support vector machine algorithm that classifies data into simple units and examines how similar instances are in a dataset |

the main data sources managed by companies that work with DS in DM.

In DS, databases are made up of different variables or indicators. These databases are known as "datasets" or "data records" (thereafter, the term "dataset" will be used). Each of the variables contained in the datasets denotes a specific characteristic (see Table 3). Datasets can contain (i) structured or (ii) unstructured data. Structured data can be stored in tables, with each table having the same structure or attributes. By contrast, unstructured data have their own internal structure and, consequently, attributes can be organized in different ways in each table.

To provide an idea of the organization of datasets used by marketers applying DS, Table 3 presents the main characteristics of the indicators typically used in such datasets.

Attributes sometimes come from raw data. Raw data include different types of data that can offer insights to solve a problem. According to Kitchin (2014), raw data can be divided into (i) capture data and (ii) exhaust data. Capture data include measurements and observations that have been designed to collect data. By contrast, exhaust data do not include such measurements and observations and must be structured based on the problem to be solved. In previous research, case studies on metadata—i.e., data that contain files uploaded to the Internet—are analyzed by researchers as exhaust data (Janssen et al., 2020).

For dataset analysis, DS rely on the models based on Machine Learning (ML). The core of modern DS, ML provides algorithms to automatically analyze large datasets. These models can be trained by researchers (also non-technical researchers) or marketers to extract actionable insights and identify patterns. A wide array of algorithms is available. These algorithms can be used and trained by connecting to companies or researchers' databases (Dwork & Roth, 2014).

ML has evolved into what is known as Deep Learning (DL), a technology that has allowed us to change how computers process language and images. DL consists of a set of neural network models with multiple layers and units on the same network. DL is the newest form of ML; however, it is not the only one used in DS (Lies, 2019). There are different other approaches to the data using ML (see Table 4 for a summary).

Next, within the ML area, there are two main types of analysis approaches: (i) Supervised Learning (SL) and (ii) Unsupervised Learning (UL) (see Table 5). SL involves training a set of samples, including pieces of text, User Generated Content (UGC), such as tweets or Facebook posts, feelings about a product, and so forth. All these samples can be used to train an algorithm. Once the expected success rate (accuracy) of the algorithm is achieved, the algorithm that works with ML can automatically perform the analysis automatically.

Unlike SL, Unsupervised Learning automatically identifies patterns that have not yet been detected in the data. In this case, human supervision is minimal. The algorithms that work with ML are called as Support Vector Machines (SVM), known as one-class classifiers. An SVM examines the dataset to identify the main characteristics and similar behaviors of the instances that make up the database and can be trained on a recurring basis. SVM will classify the dissimilar values of the instances so that the investigator can study the identified anomalies.

Data analysis processes should be framed using relevant concepts. Accordingly, in the DS field, the concepts of Data Mining (DMI) and Knowledge Discovery (KD) have been introduced. At present, these two concepts (DMI and KW) are used indiscriminately by researchers to refer to dataset analysis strategies. The concept of DMI, proposed by Lovell and Michael (1983), initially emerged in the business world to make sense of the datasets developed in data warehouses. Accordingly, today, the concept of DMI is more used in the world of business and marketing to refer to processes of discovery and identification of patterns in datasets. By contrast, KD stems from the concept of Knowledge Discovery in Databases (KDD) (Shapiro, 1989), a concept that is now more extensively used in the scientific world. It is a more technical concept to refer to processes of discovery and identification of patterns in datasets (Rudder, 2014).

**Table 6**
Search terms used in the SLR.

| Search terms | | | Data bases | Fields |
|---|---|---|---|---|
| Data sciences | AND | digital marketing | ACM Digital Library | Title |
| OR data mining* | | OR online marketing* | AIS Electronic Library | Abstract |
| OR knowledge* discovery | | OR internet marketing* | IEEE Explore | Keywords |
| | | | ScienceDirect Web of Sciences | |

* These terms were only used when the search of the terms "Data Sciences and Digital Marketing" did not yield the expected results.

## Methodology development

In this study, in order to address the research questions formulated in the previous section, the methodology of a Systematic Literature Review (SLR) was used. SLP is defined as a method that enables tackling "*an emerging issue that would benefit from exposure to potential theoretical foundations*" (Stieglitz, Mirbabaie, Ross, & Neuberger, 2018; Webster & Watson, 2002). As discussed previously, the emerging area of DS applied to the DM sector will benefit from a logical conceptualization of the application of DS to this digital environment.

For the development of SLR, the following three steps are usually used (Stieglitz et al., 2018). First, the theoretical foundations of a framework are presented for the classification of the main DS concepts used in the literature. In doing so, we follow Bem (1995) who argued that a coherent review requires a conceptual coherent structuring of the topic itself. Therefore, the present review focused on the analysis methods, uses, and performance metrics for the use of DS in DM. The main contributions of relevant studies were identified and categorized in terms of their priority for the theoretical framework.

In the second step, according to Stieglitz et al. (2018), the literature is systematically examined to identify similarities and details of the DM sector in DS. This step is used to inductively synthesize prior research and group basic concepts and definitions (Devece, Ribeiro-Soriano, & Palacios-Marqués, 2019). In the third step, the main findings of the analysis of DS in DM in the literature are considered, highlighting the main uses, applications, indicators and techniques, as well as outlining directions of further research in this field.

In the present study, we followed the guidelines proposed by vom Brocke et al. (2009, 2015) and Stieglitz et al. (2018). First, the predefined and selected terms were searched in the databases regarding title, abstract, and keywords. The irrelevant results were eliminated. The search terms were chosen to identify the main uses, applications, indicators and techniques, as well as the future of DS in DM according to the theoretical framework (see Table 6). The results were classified by categories and filters related to CS, IS, DM, marketing and business. Furthermore, only original articles and reviews were analyzed. Proceedings, books, chapters or magazines were not included in the SLR process.

This study is database-oriented and took into account all articles published and indexed in the following scientific databases: ACM Digital Library, AIS Electronic Library, IEEE Explore, ScienceDirect, and Web of Science. The detailed list of search is shown in Table 6.

Second, to identify the potential of the found articles, titles, abstracts, and keywords were read in detail. For our review, relevant research studies were those that identified the main analysis methods, uses, performance metrics and future of DS research in DM; therefore, the type of analysis and methodologies used in the studies were not taken into account when selecting the articles.

**Table 7**
SLR results.

| Database | Number of results | Number of relevant results |
|---|---|---|
| ACM Digital Library | 136 | 13 |
| AIS Electronic Library | 5 | 2 |
| IEEE Explore | 79 | 10 |
| ScienceDirect | 34 | 9 |
| Web of Sciences | 105 | 15 |
| Total | 354 | 49 |

**Table 8**
Number of article classifications by results.

| Article classification | Data Sciences | Digital Marketing | Mixed Articles |
|---|---|---|---|
| Classification | 16 | 13 | 20 |
| Methods | 12 | 4 | 11 |
| Uses | 6 | 6 | 11 |
| Performance metrics | 4 | 11 | 11 |
| Future topics | 4 | 5 | 9 |

Thirdly, selected articles were categorized as relevant following the definitions, applications, and theoretical concepts regarding the importance of applying DS techniques in DM. Therefore, the articles that contained inadequate terms or were not conclusive, did not contain the search terms, had no relation to the research topic, offered no quality evaluation, or contained no description and specification of terms were removed (see Fig. 1).

The steps of the development of the SLR undertaken in the present study, performed following vom Brocke et al. (2015) and Stieglitz et al. (2018) and enriched by the SLR presented by the PRISMA diagram (Moher, Liberati, Tetzlaff, Altman, & T, 2009), are shown in Fig. 1.

Table 7 shows the total number of articles identified based on each of the objectives proposed in the SLR process.

While the very nature of the SLR indicates that the results from the databases should be related to both DS and DM, the articles were classified from the point of view of the factors analyzed. That is, if some articles were classified as having the main focus of DS, the analyzed factors were justified from the DS point of view. On the other hand, for those articles classified as focusing on DM, the concepts analyzed were from the point of view of DM and its influence on DS. Furthermore, there were articles analyzed from a broader point of view as they focused on methods, uses, performance metrics and future topics for both categories. In addition, articles that presented different categories (methods, uses, performance metrics and future topics) within the same article were also analyzed (see Table 8).

Table 9 summarizes all studies included in the review, with the specifications of authors, journals, research category, and content (specifically, main topic and main focus).

## Analysis of results

DS provide different perspectives and approaches to statistical data analysis. Statistics is a set of rules to quantitatively analyze any type of data. However, with the evolution of mathematics and the development of DS, statistical learning has been defined as a theoretical framework that works with ML from the point of view of DS (Tsapatsoulis & Djouvas, 2019). Therefore, the objective of the methods used in DS applying statistical learning is to perform (i) functional analysis, (ii) exploratory analysis, and (iii) prediction of results based on the analyzed data. Table 10 provides a summary of the main methods identified in and applied to the DM ecosystem.

After the analysis of the main methods applied to the study of DM using the DS techniques, we identified the uses and applications

**Table 9**
Relevant studies found in the Systematic Literature Review.

| Authors | Journal | Category | Main topic | | Main focus | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Data Sciences | Digital Marketing | Methods | Uses | Perform. Metric | Future Topics |
| Alrifai (2017) | The Journal of Computing Sciences in Colleges | Computer Sciences | ● | | ● | ● | ● | |
| Arias, Arratia, and Xuriguera (2014) | ACM Transactions on Intelligent Systems and Technology | Intelli. Sys. & Tech. | ● | ● | ● | | ● | |
| Ballestar, Grau-Carles, and Sainz (2018) | Journal of Business Research | Business | | ● | | ● | ● | |
| Beaput, Banik and Joshi (2019) | The Journal of Computing Sciences in Colleges | Computer Sciences | ● | | | | | ● |
| Cassavia, Masciari, Pulice, and Sacca (2017) | ACM Transactions on Interactive Intelligent Systems | Interactive Intelli. Sys. | ● | ● | ● | | ● | |
| Cheng and Wang (2018) | Journal of Business Economics and Management | Business, Economics | | ● | ● | | ● | ● |
| Dadzie, Sibarani, Novalija, and Scerri (2018) | Journal of Web Semantics | Information Systems | ● | | ● | | | |
| De Caigny, Coussement, and De Bock (2020) | Decision Support Systems | Computer Sciences | ● | ● | ● | ● | | ● |
| Debortoli, Müller, and vom Brocke (2014) | Business & Information Systems Engineering | Business, Info. Systems | ● | ● | | ● | ● | ● |
| Dwivedi et al. (2019) | International Journal of Information Management | Information Sciences | ● | | | ● | | ● |
| Fan et al. (2015) | Proceedings of the VLDB Endowment | Computer Sciences | ● | | ● | ● | | |
| Guinan, Parise, and Langowitz (2019) | Business Horizons | Business | | ● | | ● | | ● |
| Haq, Li, and Hassan (2019) | IEEE Access | Computer Sciences | ● | | ● | ● | | |
| Hodge et al. (2018) | IEEE Transactions on Games | Computer Sciences | ● | | | ● | | ● |
| Jacobson, Gruzd, and Hernández-García (2020) | Journal of Retailing and Consumer Services | Business | ● | ● | | ● | | ● |
| Jiao, Wang, Feng, and Niyato (2018) | IEEE Internet of Things Journal | Computer Sciences | ● | | ● | | | ● |
| Lagrée, Cappé, Cautis, and Maniu (2018) | ACM Transactions on Knowledge Discovery from Data | Information Science | | ● | | ● | ● | |
| Lau et al. (2012) | ACM Transactions on Management Information Systems | Information Systems | ● | ● | ● | ● | | |
| Liu et al. (2016) | ACM Transactions on Knowledge Discovery from Data | Information Sciences | | ● | ● | ● | ● | |
| Liu, Gu, Ko, and Liu (2018) | IEEE transactions on cybernetics | Automa. & Control Syst. | ● | ● | | ● | ● | |
| Miklosik, Kuchta, Evans, and Zak (2019) | IEEE Access | Computer Sciences | ● | ● | | ● | ● | |
| Milovanović, Bogdanović, Labus, Barać, and Despotović-Zrakić (2019) | Future Generation Computer Systems | Computer Sciences | ● | | ● | ● | | |
| Naqvi, Awais, Saeed, and Ashraf (2018) | Inter. J. of Advanced Computer Science and Applications | Com. Sc.,The. & Metho. | ● | | ● | | ● | |
| Papoutsoglou, Ampatzoglou, Mittas, and Angelis (2019) | IEEE Access | Computer Sciences | ● | | ● | | | |

Table 9 (*Continued*)

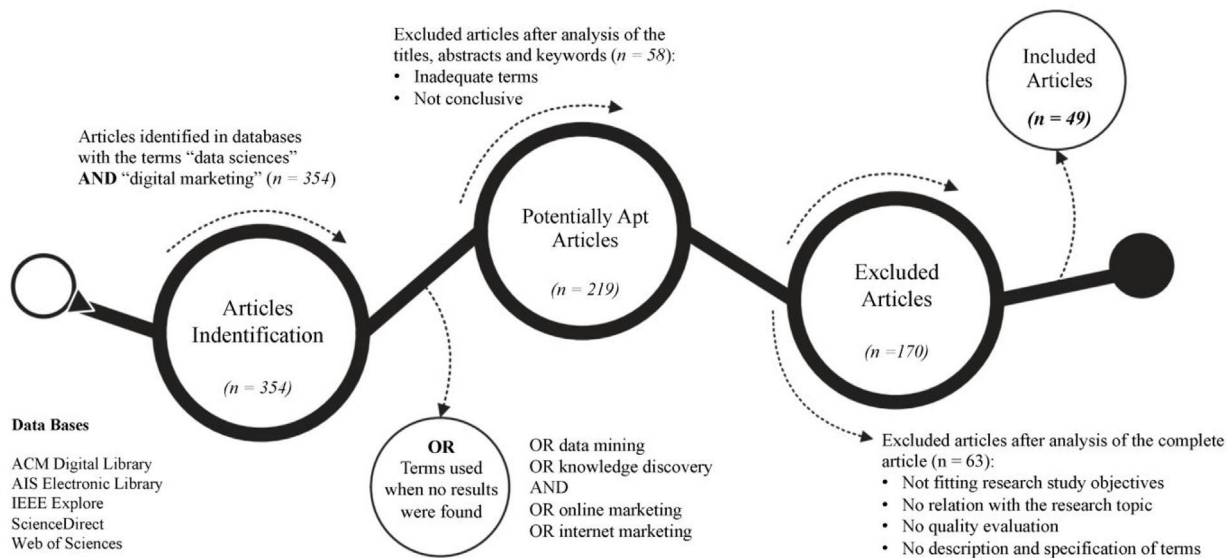| Authors | Journal | Category | Main topic | | Main focus | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Data Sciences | Digital Marketing | Methods | Uses | Perform. Metric | Future Topics |
| Poddar, Banerjee, and Sridhar (2019) | Journal of Business Research | Business | | • | | • | • | |
| Roundy and Asllani (2018) | Journal of Entrepreneurship, Management and Innovation | Business, Economics | • | • | | • | • | • |
| Ruggieri, Pedreschi, and Turini (2010) | ACM Transactions on Knowledge Discovery from Data | Information Sciences | • | | • | | • | |
| Saleemi, Anjum, and Rehman (2017) | IEEE Access | Computer Sciences | • | | • | • | | • |
| Salminen, Yoganathan, Corporan, Jansen, and Jung (2019) | Journal of Business Research | Business | • | • | • | | • | • |
| Sang and Xu (2011) | ACM Trans. Mult. Comp., Commu., and Applications | Computer Science | • | • | • | | • | • |
| Saura and Bennett (2019) | Symmetry-Basel | Multidisciplinary | • | • | • | • | • | |
| Saura, Herráez, and Reyes-Menendez (2019) | IEEE Access | Computer Sciences | • | • | • | | • | |
| Spiekermann, Korunovska, and Langheinrich (2018) | Proceedings of the IEEE | Engineering | • | | | | | • |
| Troisi, Maione, Grimaldi, and Loia (2019) | Industrial Marketing Management | Business, Management | | • | | | • | • |
| Tsapatsoulis and Djouvas (2019) | Frontiers in Robotics and AI | Robotics | | • | • | | • | |
| Wu and Holsapple (2013) | ACM Transactions on Knowledge Discovery from Data | Information Sciences | | • | | • | • | • |
| Wymbs (2016) | Journal of Information Systems Education | Information Systems | | • | | • | • | |
| Yeo, Hwang, Kim, Koh, and Lipka (2018) | IEEE Transactions on Knowledge and Data Engineering | Information Sciences | • | • | • | • | | |
| Zannettou, Sirivianos, Blackburn, and Kourtellis (2019) | Journal of Data and Information Quality | Data and Infor. Quality | • | • | | • | • | |
| Zhu, Peng, Chen, Zheng, and Zhou (2016) | ACM Transactions on Knowledge Discovery from Data | Information Sciences | | • | | • | • | • |
| Wang, Zhang, and Yuan (2017) | Foundations and Trends® in Information Retrieval | Computer Sciences | • | • | | | • | • |
| Song et al. (2014) | Personal and ubiquitous computing | Computer Sciences | • | • | • | • | | |
| Chang, Kauffman, and Kwon (2014) | Decision Support Systems | Computer Sciences | • | | | | | • |
| Kaiser and Bodendorf (2012) | Internet Research: Electronic Netw. Applic. and Policy | Business & Info. Sci. | | • | | • | • | |
| Archak et al. (2011) | Management science | Management | • | • | | • | | • |
| Liao, Chen, Hsieh, and Hsiao (2009) | Expert Systems with Application | Computer Sciences | • | • | • | | • | • |
| Su, Chen, and Sha (2007) | International Journal of Technology Management | Management | | • | • | • | | • |
| Su, Chen, and Sha (2006) | Technovation | Engineering & Managem. | • | | • | | • | |
| Kwan, Fong, and Wong (2005) | Decision Support Systems | Computer Sciences | • | • | | • | • | • |

**Fig. 1.** PRISMA flow diagram of SLR.

**Table 10**
Type of methods used in Data Sciences as applied to Digital Marketing.

| Type of analysis | Description |
| --- | --- |
| Descriptive statistics | Descriptive statistics is used to quantitatively summarize features from a collection of information or a dataset. It includes the measurement of central tendency, arithmetic mean, or measures such as variance or the range |
| Bayes' rule | Bayer's rule is used for the description of probabilities of events. It is based on knowledge about the conditions that could have caused a specific event |
| Method of least squares | The method of least squares makes it possible to find the best theoretical model made up of variables or constructs that fit a dataset and allows for a quantitative validation |
| Linear regression | Linear regression is used to model the relationship between a scale and an exploratory variable. When there is more than one analysis variable within the linear regression, it is called multiple linear regression |
| Logistic regression | Logistic regression is a regression analysis used to predict a categorical variable. A categorical variable is a variable that can be subdivided into more categories. It is used as a general rule to model the probability of an event and the factors that compose it |
| Artificial neural network | Artificial neural networks are self-learning systems made up of interconnected neurons of nodes that have an input and an output. They are used to find or detect solutions or features that are otherwise difficult to identify using standard programming |
| Multivariate analysis | Multivariate analysis is an analysis model focused on the multiple analysis of data collected from more than one dependent variable. Multivariate analysis involves the analysis of variables and correlations among them |
| Maximum likelihood estimate | Maximum likelihood estimate is a method for estimating statistical parameters in an observation. Its results show the highest probability for the estimation of the analyzed parameters |
| Discriminatory analysis | Discriminatory analysis is known as classification or pattern-recognition (nearest-neighbor models). It automatically recognizes patterns or regularities in a dataset and is widely used in research focused on DM or KDD |
| Information theory | Information theory is used to analyze and study the quantification and communication of information to find the limits of communication processing |
| Artificial Intelligence | Artificial Intelligence is focused on the simulation of human intelligence by machines. In the present study, Artificial Intelligence refers to using machines automatically focused on machine learning and problem solving |

pursued by the DS in the tactical strategies and developments of the DM (see Table 11).

In the DM ecosystem, one of the main challenges is controlling and defining the success of a DS strategy. To this end, marketers and researches should choose and understand the main performance metrics for the measurement of the models and methodologies used. The results of our SLR analysis highlighted the following indicators to measure the success of these SD strategies applied to the DM sector (see Table 12).

Fig. 2 shows the main topics DM research using DS. These research areas were selected taking into account the recommendations of the studies analyzed and their discussions of further research in the field.

In what follows, each of the topics and the influence that these topics may have on the development of DM strategies using DS are explained in detail.

*Medical data and eHealth strategies:* Analysis of users' medical data can help to find trends and thus facilitate the creation of new vaccines, fighting diseases that can cause an uncontrolled epidemic (such as the coronavirus known as COVID-19), and predicting possible deaths. Marketing must promote companies' use of such strategies to collect data for further analysis.

*Smart cities & governance:* Efficient management of energy resources, as well as sustainable and intelligent construction and development are based on automation and artificial intelligence of large structures (Ismagilova, Hughes, Dwivedi, & Raman, 2019; Janssen, Luthra, Mangla, Rana, & Dwivedi, 2019). Social or responsible marketing, also known as corporate social responsibility (CSR), is driven by DM-based communications through digital platforms and social media channels (Orlandi, Ricciardi, Rossignoli, & De Marco, 2019).

*Internet of Things (IoT):* IoT refers to management and collection of daily use data from connected devices. This also includes order and identification of new features that help personalize and offer new products and services and to create new needs (Brous, Janssen, & Herder, 2020). DM adapts to the mobile environment

**Table 11**
Mean uses of Data Sciences in Digital Marketing strategies.

| Type | Description |
|---|---|
| Analyzing User Generated Content (UGC) | Analysis of the content publicly published by users on social networks and digital platforms. Some examples include tweets, reviews, reviews, comments, shares, likes, meta data, etc |
| Optimize customers' preferences | Optimization of user preferences in products or services customizable based on the analysis of large databases containing information about customer preferences |
| Tracking customer behavior online | Tracking and predicting consumer behavior on digital channels. In this way, companies can anticipate their marketing and advertising actions and better understand users |
| Tracking social media commentary/Interactions | Tracking user-company responses and interactions in digital ecosystems. Based on the results, marketers can transmit more relevant messages |
| Optimize stock levels in ecommerce websites | Optimization of the stock in e-commerce websites for better management of both the information and the products and services to anticipate sales and request products from suppliers in advance |
| Analysis of online sales data | Sales analysis to find patterns that explain trends, current demand and customers' interest in specific products |
| Introducing new products | Market research to find implausible patterns that can offer new products to sectors saturated with offers |
| Analyzing social media trends | Analysis of the trends in social media to avoid, for example, crisis of reputation of companies or social movements that require attention by the company. Also known as Social Listening |
| Analyzing product recommendations and reviews | Analysis of consumer reviews and recommendations to identify key points and characteristics of companies' products, services, or customer service |
| Personalizing customer's online experience | Improving customer shopping experiences by offering personalized experiences based on information from previous users who visited a website |
| Building recommender systems | Creation of automatic recommendation systems that establish priorities and predictions of purchase success for certain customers based on the analysis of their purchase history |
| Measure and predict clicks online (Social and Paid Ads) | Prediction of user clicks on web pages and social networks. Based on this information, companies can improve the profitability of their ads and increase the visibility of impressions and clicks, also known as Click Through Rate (CTR) |
| Measure and predict user's behavior | Predicting how users will behave to offer or avoid problems before a purchase, e.g. sending emails in email marketing campaigns to offer additional products and thus increase profitability |
| Improve User Experience (UX) | Improving user experience in terms of graphic and visual design of web pages, mobile applications, etc. focused on the "user-friendly" movement |
| Analyze real-time data | Facilitation of direct decisions regarding certain situations of sales, visibility, or dissemination of information |
| Identify online communities | Identification of online communities and their influential and leaders. Analysis of the user considerations and loyalty to opinion leaders and their messages |
| Identificar fake news y contenidos falsos | Identification of false content about companies or shared advertising messages; alternative terms are "fakenews" or "deepfakes" |

**Table 12**
Mean performance metrics to measure the success of DS approaches in DM.

| Indicator | Description |
|---|---|
| Reliability | Reliability is the metric of accuracy and exhaustiveness of the processing of a dataset linked to its use |
| Accuracy | Accuracy, also referred to as precision, shows the quality and accuracy of hit of a model or method |
| Precision | Precision, also known as Positive Predictive Value (PPV), is a metric that measures the relevance and success of the approach of a method within the database where DS has been applied |
| Validity | Validity is the measure that indicates whether the data support the results and conclusions obtained in a truthful and accurate way. |
| Consistency | Consistency evaluates whether the values represented in data from one dataset are consistent with the values represented in the data from another dataset at the same point in time |
| Recall | Recall, also called sensitivity, generally refers to the number of correct results divided by the number of discarded values |
| Sensitivity | Sensitivity, sometimes referred to as probability of detection or positive rate, measures the proportion of positive values or points identified in a dataset |
| Specificity | Specificity is the metric that evaluates the prediction characteristics of true negatives in the variables within a category in a dataset |
| Prevalence | Prevalence refers to the proportion of population with specific common characteristics during a specific period of time |

with mobile-friendly design initiatives and strategies that fosu on connected devices.

*Data privacy and management: Mass data management:* This includes rights, access, and legitimate profitability of large public databases (Nissenbaum, 2009). One of the functions of DM is to raise consumer awareness about how companies will make use of their data (emails, phones, demographics, and so forth) (Lee & Trimi, 2018).

*People: movement, organization and personalization:* This includes analyzing the movement and organization of people through the analysis of large databases of citizens or vehicle purchases (Aladwani & Dwivedi, 2018; Höflinger, Nagel, & Sandner, 2018). The DM has the challenge of personalizing massive messages and, using the DS methods, identifying specific habits according to the type of people and their demographic and psychographic characteristics to increase the ROI of digital campaigns (Palacios-Marqués, García, Sánchez, & Mari, 2019).

*Development of new Machine Learning models:* This includes new Machine Learning models that companies can train and apply in their projects. There is a growing need for the creation of models focused on solving specific problems. These models should be created, trained, and debugged for a specific purpose. Other tasks include designing user-friendly algorithms and ML models to break the technical barrier between marketers and data specialists (Caseiro & Coelho, 2019).

*Operational CRM and data management:* This includes the creation of automatic company information management systems that can identify better unsuspected patterns and extract actionable insights to help company to manage their information in real time and to enable marketing experts to take better decisions (Ricciardi, Zardini, & Rossignoli, 2018).

*Sustainable strategies based on data:* This includes the study of the sources of data resources and management of globalization processes to increase sustainable strategies and actions based on data analysis. Relevant research areas in this field include social marketing or green marketing (Archak, Ghose, & Ipeirotis, 2011).

*Social media listening:* This includes automated research on important trends in social networks and messages released by opinion leaders, as well as exploring the responses of communities to massive messages in the face of crises, epidemics, as well
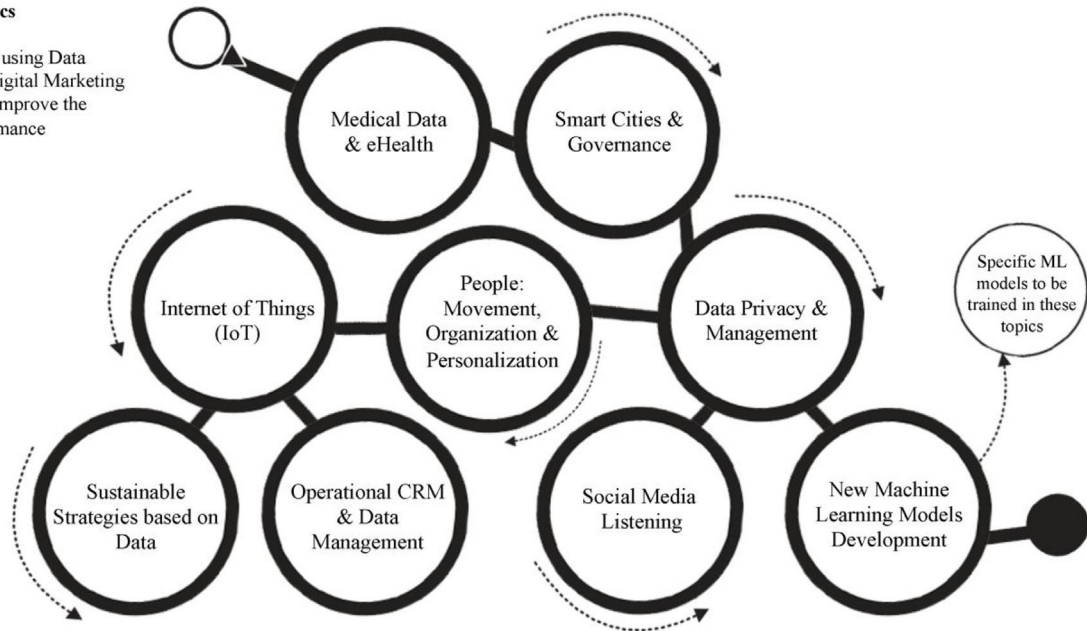
**Fig. 2.** Main topics of research in Digital Marketing using Data Sciences.

as environmental or social movements (Reyes-Menendez, Saura, & Stephen, 2020). DM should understand how these communities are organized and take adequate action with persuasive and responsible messages.

## Conclusions

In this review article, we have defined the main concepts, methods, and performance metrics used in DS throughout the last two decades and their applications in DM. We have provided a structured account of the main concepts that marketers should take into account when considering a DM strategy based on data intelligence. Pertinent methods used in DS to extract actionable insights from large amounts of data have also been identified. We have also outlined major performances metrics used to measure the DS performance in the DM environment. These results respond to first research question addressed in the present study (*What are the main methods of analysis, uses, and performance metrics of Data Sciences applied in Digital Marketing?*).

Today, companies are involved in an increasingly data-driven ecosystem. Accordingly, the number of ML-based user-friendly applications that companies, marketers, and non-technical researchers can use has considerably increased. However, the understanding by marketers and marketing researchers of the main notions of DS is essential to be efficient and lasting over time, as the lack of such understanding has already become a skill problem (Ghotbifar et al., 2017; Royle & Laing, 2014).

Specifically, businesses have been reported to waste a lot of time organizing, cleaning, and structuring the databases of their users and customers (Kelleher & Tierney, 2018). In this context, the use of relevant indicators and performance metrics will help companies, marketers, and non-technical researchers in the marketing area to conduct better research and to more efficiently measure the time they spend analyzing and structuring their databases.

With regard to our second research question ("*What are the areas of further research on the use of Data Science in Digital Marketing?*"), in our results, we have identified a total of 9 topics for future research on DS in the DM ecosystem. Undoubtedly, the application of new

specific ML models to each of these topics will define the future of the sector in terms of the effectiveness of its data-driven strategies.

As for the theoretical implications, the present review has identified a total of 11 methods, 17 uses, 9 performance metrics, and 9 research topics that can be used by researchers as the starting points of their research focused on the use of DS in strategies of DM. Regarding the methods, when considering their DM research using DS analysis, non-technical researchers can take into account which of these models best fit the objectives of their research based on the definitions presented. Furthermore, for the elaboration of new studies, researchers in the DM sector can use the nine identified topics to formulate new hypotheses or to find new research questions that need be addressed.

Furthermore, the present review offers important practical implications for the industry. Today, companies are increasingly developing data-driven strategies. Therefore, the best use of these strategies requires an in-depth understanding of all necessary steps. The results of the present review can be meaningfully used to familiarize companies' experts with the main DS indicators and metrics in the DM ecosystem.

Consequently, companies can use the results of the present review as the starting point in the elaboration of new DM strategies. Companies can consider implementing any of the 17 identified uses of DS in DM to obtain actionable insights from their datasets. In addition, in terms of performance metrics, companies can take the definitions and descriptions provided in this review to make reports and present their contingency and control plans, as well as to measure the success of their digital campaigns.

The limitations of this study include the number of databases analyzed and the criteria used to collect the articles from the databases. In addition, the selection of the articles and their classification could have biased the final results. Further research should focus on the topics similar to the 9 topics identified in the present review, In addition, the improvement of DS processes in the DM ecosystem, which has not been identified in this review, should also be considered.

# References

Aladwani, A. M., & Dwivedi, Y. K. (2018). Towards a theory of SocioCitizenry: Quality anticipation, trust configuration, and approved adaptation of governmental social media. *International Journal of Information Management*, *43*, 261–272.

Alrifai, R. (2017). A data mining approach to evaluate stock-picking strategies. *Journal of Computing Sciences in Colleges*, *32*(5), 148–155.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509.

Arias, M., Arratia, A., & Xuriguera, R. (2014). Forecasting with twitter data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(1), 1–24.

Avery, J., Steenburgh, T. J., Deighton, J., & Caravella, M. (2012). Adding bricks to clicks: Predicting the patterns of cross-channel elasticities over time. *Journal of Marketing*, *76*(3), 96–111.

Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research*, *88*, 407–414.

Beaput, A., Banik, S., & Joshi, D. (2019). Ranking privacy of the users in the cyberspace. *The Journal of Computing Sciences in Colleges*, *35*(4), 109–114.

Bem, D. J. (1995). Writing a review article for Psychological Bulletin. *Psychological Bulletin*, *118*(2), 172–177. http://dx.doi.org/10.1037/0033-2909.118.2.172

Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Hoboken, New Jersey: John Wiley & Sons.

Braverman, S. (2015). Global review of data-driven marketing and advertising. *Journal of Direct Data and Digital Marketing Practice*, *16*(3), 181–183.

Brous, P., Janssen, M., & Herder, P. (2020). The dual effects of the Internet of Things (IoT): A systematic review of the benefits and risks of IoT adoption by organizations. *International Journal of Information Management*, *51*, 101952.

Caseiro, N., & Coelho, A. (2019). The influence of Business Intelligence capacity, network learning and innovativeness on startups performance. *Journal of Innovation & Knowledge*, *4*(3), 139–145. http://dx.doi.org/10.1016/j.jik.2018.03.009

Cassavia, N., Masciari, E., Pulice, C., & Sacca, D. (2017). Discovering user behavioral features to enhance information search on big data. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*(2), 1–33.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*, 67–80.

Cheng, F. C., & Wang, Y. S. (2018). The do not track mechanism for digital footprint privacy protection in marketing applications. *Journal of Business Economics and Management*, *19*(2), 253–267.

Codd, E. F. (1971). A data base sublanguage founded on the relational calculus. In *Proceedings of the 1971 ACM SIGFIDET (now SIGMOD) workshop on data description, access and control* (pp. 35–68). November.

Dadzie, A. S., Sibarani, E. M., Novalija, I., & Scerri, S. (2018). Structuring visual exploratory analysis of skill demand. *Journal of Web Semantics*, *49*, 51–70.

Davenport, T. (2014). *Big data at work: dispelling the myths, uncovering the opportunities*. Brighton, Massachusetts: Harvard Business Review Press.

De Caigny, A., Coussement, K., & De Bock, K. W. (2020). Leveraging fine-grained transaction data for customer life event predictions. *Decision Support Systems*, *130*, 113232.

Debortoli, S., Müller, O., & vom Brocke, J. (2014). Comparing business intelligence and big data skills. *Business & Information Systems Engineering*, *6*(5), 289–300.

Devece, C., Ribeiro-Soriano, D. E., & Palacios-Marqués, D. (2019). Coopetition as the new trend in inter-firm alliances: literature review and research patterns. *Review of Managerial Science*, *13*(2), 207–226.

Dremel, C., Herterich, M. M., Wulf, J., & Vom Brocke, J. (2020). Actualizing big data analytics affordances: A revelatory case study. *Information & Management*, *57*(1), 103121.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., et al. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3–4), 211–407.

Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research*, *2*(1), 28–32.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137–144.

Ghotbifar, F., Marjani, M., & Ramazani, A. (2017). Identifying and assessing the factors affecting skill gap in digital marketing in communication industry companies. *Independent Journal of Management & Production*, *8*(1), 1–14.

Guinan, P. J., Parise, S., & Langowitz, N. (2019). Creating an innovative digital project team: Levers to enable digital transformation. *Business Horizons*, *62*(6), 717–727.

Haq, M. I. U., Li, Q., & Hassan, S. (2019). Text mining techniques to capture facts for cloud computing adoption and big data processing. *IEEE Access*, *7*, 162254–162267.

Hodge, V., Sephton, N., Devlin, S., Cowling, P., Goumagias, N., Shao, J., et al. (2018). How the business model of customisable card games influences player engagement. *IEEE Transactions on Games*, *11*(4), 374–385.

Höflinger, P. J., Nagel, C., & Sandner, P. (2018). Reputation for technological innovation: Does it actually cohere with innovative activity? *Journal of Innovation & Knowledge*, *3*(1), 26–39. http://dx.doi.org/10.1016/j.jik.2017.08.002

Ismagilova, E., Hughes, L., Dwivedi, Y. K., & Raman, K. R. (2019). Smart cities: Advances in research – An information systems perspective. *International Journal of Information Management*, *47*, 88–100.

Jacobson, J., Gruzd, A., & Hernández-García, Á. (2020). Social media marketing: Who is watching the watchers? *Journal of Retailing and Consumer Services*, *53*, 53.

Janssen, M., Luthra, S., Mangla, S., Rana, N. P., & Dwivedi, Y. K. (2019). Challenges for adopting and implementing IoT in smart cities. *Internet Research*, *29*(6), 1589–1616.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, 101493.

Jiao, Y., Wang, P., Feng, S., & Niyato, D. (2018). Profit maximization mechanism and data management for data analytics services. *IEEE Internet of Things Journal*, *5*(3), 2001–2014.

Kaiser, C., & Bodendorf, F. (2012). Mining consumer dialog in online forums. *Internet Research: Electronic Networking Applications and Policy*, *22*(3), 275–297.

Kannan, P. K. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, *34*(1), 22–45.

Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.

Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Cambridge, Massachusetts: MIT Press.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Newbury Park, California: Sage.

Kumar, V., Chattaraman, V., Neghina, C., Skiera, B., Aksoy, L., Buoye, A., et al. (2013). Data-driven services marketing in a connected world. *Journal of Service Management*, *24*(3), 330–352.

Kwan, I. S., Fong, J., & Wong, H. K. (2005). An e-customer behavior model with online analytical mining for internet marketing planning. *Decision Support Systems*, *41*(1), 189–204.

Lagrée, P., Cappé, O., Cautis, B., & Maniu, S. (2018). Algorithms for online influencer marketing. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *13*(1), 1–30.

Lau, R. Y., Liao, S. Y., Kwok, R. C. W., Xu, K., Xia, Y., & Li, Y. (2012). Text mining and probabilistic language modeling for online review spam detection. *ACM Transactions on Management Information Systems (TMIS)*, *2*(4), 1–30.

Lee, S. M., & Trimi, S. (2018). Innovation for creating a smart future. *Journal of Innovation & Knowledge*, *3*(1), 1–8. http://dx.doi.org/10.1016/j.jik.2016.11.001

Leeflang, P. S., Verhoef, P. C., Dahlström, P., & Freundt, T. (2014). Challenges and solutions for marketing in a digital era. *European Management Journal*, *32*(1), 1–12.

Liao, S. H., Chen, C. M., Hsieh, C. L., & Hsiao, S. C. (2009). Mining information users' knowledge for one-to-one marketing on information appliance. *Expert Systems with Applications*, *36*(3), 4967–4979.

Lies, J. (2019). Marketing intelligence and big data: digital marketing techniques on their way to becoming social engineering techniques in marketing. *International Journal of Interactive Multimedia & Artificial Intelligence*, *5*(5).

Liu, B., Wu, Y., Gong, N. Z., Wu, J., Xiong, H., & Ester, M. (2016). Structural analysis of user choices for mobile app recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *11*(2), 1–23.

Liu, Y., Gu, Z., Ko, T. H., & Liu, J. (2018). Identifying key opinion leaders in social media via modality-consistent harmonized discriminant embedding. *IEEE Transactions on Cybernetics*, *50*(2), 717–728.

Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, *24*(3), 149–157.

Lovell, & Michael, C. (1983). "Data Mining". *The Review of Economics and Statistics*, *65*(1), 1–12. http://dx.doi.org/10.2307/1924403

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, Massachusetts: Houghton Mifflin Harcourt>.

Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the adoption of machine learning-based analytical tools in digital marketing. *IEEE Access*, *7*, 85705–85718.

Milovanović, S., Bogdanović, Z., Labus, A., Barać, D., & Despotović-Zrakić, M. (2019). An approach to identify user preferences based on social network analysis. *Future Generation Computer Systems*, *93*, 121–129.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and MetaAnalyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097. http://dx.doi.org/10.1371/journal.pmed1000097

Naqvi, S. M. A., Awais, M., Saeed, M. Y., & Ashraf, M. M. (2018). Opinion mining and thought pattern classification with natural language processing (NLP) tools. *International Journal of Advanced Computer Science and Applications*, *9*(10), 485–493.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Redwood City, California: Stanford University Press.

Orlandi, L. B., Ricciardi, F., Rossignoli, C., & De Marco, M. (2019). Scholarly work in the Internet age: Co-evolving technologies, institutions and workflows. *Journal of Innovation & Knowledge*, *4*(1), 55–61. http://dx.doi.org/10.1016/j.jik.2017.11.001

Palacios-Marqués, D., García, M. G., Sánchez, M. M., & Mari, M. P. A. (2019). Social entrepreneurship and organizational performance: A study of the mediating role of distinctive competencies in marketing. *Journal of Business Research*, *101*, 426–432.

Palacios-Marqués, D., Welsh, D. H., Méndez-Picazo, M. T., Palacios-Marqués, D., Welsh, D. H. B., & Méndez-Picazo, M. T. (2016). The importance of the activities of innovation and knowledge in the economy: Welcome to the Journal of Innovation and Knowledge. *Journal of Innovation and Knowledge, 1*(1), 1–2.

Palos-Sanchez, P., Saura, J. R., & Martin-Velicia, F. (2019). A study of the effects of programmatic advertising on users' concerns about privacy overtime. *Journal of Business Research, 96*, 61–72.

Papoutsoglou, M., Ampatzoglou, A., Mittas, N., & Angelis, L. (2019). Extracting knowledge from on-line sources for software engineering labor market: A mapping study. *IEEE Access, 7*, 157595–157613.

Piatetsky-Shapiro, G.;1; (1989). Notes of IJCAI'89 Workshop Knowledge Discovery in Databases KDD'89. Detroit, Michigan, July.

Poddar, A., Banerjee, S., & Sridhar, K. (2019). False advertising or slander? Using location based tweets to assess online rating-reliability. *Journal of Business Research, 99*, 390–397.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking.* Sebastopol, California: O'Reilly Media, Inc.

Reyes-Menendez, A., Saura, J. R., & Stephen, B. T. (2020). Exploring key indicators of social identity in the #MeToo era: Using discourse analysis in UGC. *International Journal of Information Management, 54*, 102129. http://dx.doi.org/10.1016/j.ijinfomgt.2020.102129

Ricciardi, F., Zardini, A., & Rossignoli, C. (2018). Organizational integration of the IT function: A key enabler of firm capabilities and performance. *Journal of Innovation & Knowledge, 3*(3), 93–107. http://dx.doi.org/10.1016/j.jik.2017.02.003

Rogers, D., & Sexton, D. (2012). Marketing ROI in the era of big data. In *The 2012 BRITENYAMA Marketing in Transition Study.* pp. 1–17.

Roundy, P. T., & Asllani, A. (2018). The themes of entrepreneurship discourse: A data analytics approach. *Journal of Entrepreneurship, Management and Innovation, 14*(3), 127–158.

Royle, J., & Laing, A. (2014). The digital marketing skills gap: Developing a digital marketer model for the communication industries. *International Journal of Information Management, 34*(2), 65–73.

Rudder, C. (2014). *Dataclysm: Who we are (when we think no one's looking).* Toronto, Canada: Random House Canada.

Ruggieri, S., Pedreschi, D., & Turini, F. (2010). Data mining for discrimination discovery. *ACM Transactions on Knowledge Discovery from Data (TKDD), 4*(2), 1–40.

Saleemi, M., Anjum, M., & Rehman, M. (2017). eServices classification, trends, and analysis: A systematic mapping study. *IEEE Access, 5*, 26104–26123.

Salminen, J., Yoganathan, V., Corporan, J., Jansen, B. J., & Jung, S. G. (2019). Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type. *Journal of Business Research, 101*, 203–217.

Sang, J., & Xu, C. (2011). Browse by chunks: Topic mining and organizing on web-scale social media. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 7*(1), 1–18.

Saura, J. R., & Bennett, D. R. (2019). A three-stage method for data text mining: Using UGC in business intelligence analysis. *Symmetry, 11*(4), 519.

Saura, J. R., Herráez, B. R., & Reyes-Menendez, A. (2019). Comparing a traditional approach for financial brand communication analysis with a big data analytics technique. *IEEE Access, 7*, 37100–37108.

Saura, J. R., Palos-Sánchez, P., & Cerdá Suárez, L. M. (2017). Understanding the digital marketing environment with KPIs and web analytics. *Future Internet, 9*(4), 76.

Shareef, M. A., Kapoor, K. K., Mukerji, B., Dwivedi, R., & Dwivedi, Y. K. (2020). Group behavior in social media: Antecedents of initial trust formation. *Computers in Human Behavior, 105*, 106225.

Song, G. Y., Cheon, Y., Lee, K., Lim, H., Chung, K. Y., & Rim, H. C. (2014). Multiple categorizations of products: Cognitive modeling of customers through social media data mining. *Personal and Ubiquitous Computing, 18*(6), 1387–1403.

Spiekermann, S., Korunovska, J., & Langheinrich, M. (2018). Inside the organization: Why privacy and security engineering is a challenge for engineers. *Proceedings of the IEEE, 107*(3), 600–615.

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management, 39*, 156–168.

Su, C. T., Chen, Y. H., & Sha, D. Y. (2006). Linking innovative product development with customer knowledge: A data-mining approach. *Technovation, 26*(7), 784–795.

Su, C. T., Chen, Y. H., & Sha, D. Y. J. (2007). Managing product and customer knowledge in innovative new product development. *International Journal of Technology Management, 39*(1–2), 105–130.

Sundsøy, P., Bjelland, J., Iqbal, A. M., & de Montjoye, Y. A. (2014). Big data-driven marketing: how machine learning outperforms marketers' gut-feeling. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 367–374). Cham: Springer.

Tiago, M. T. P. M. B., & Veríssimo, J. M. C. (2014). Digital marketing and social media: Why bother? *Business horizons, 57*(6), 703–708.

Troisi, O., Maione, G., Grimaldi, M., & Loia, F. (2019). Growth hacking: Insights on data-driven decision-making from three firms. *Industrial Marketing Management.*

Tsapatsoulis, N., & Djouvas, C. (2019). Opinion mining from social media short texts: Does collective intelligence beat deep learning? *Frontiers in Robotics and AI, 5*, 138.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. *ECIS 2009 Proceedings.*

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems, 37*(1).

Wang, J., Zhang, W., & Yuan, S. (2017). Display advertising with real-time bidding (RTB) and behavioural targeting. *Foundations and Trends® in Information Retrieval, 11*(4–5), 297–435.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii.

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing, 80*(6), 97–121.

Wu, J., & Holsapple, C. W. (2013). Does knowledge management matter? The empirical evidence from market-based valuation. *ACM Transactions on Management Information Systems (TMIS), 4*(2), 1–23.

Wymbs, C. (2016). Managing the innovation process: Infusing data analytics into the undergraduate business curriculum (lessons learned and next steps). *Journal of Information Systems Education, 27*(1), 61.

Yeo, J., Hwang, S. W., Kim, S., Koh, E., & Lipka, N. (2018). Conversion prediction from clickstream: Modeling market prediction and customer predictability. *IEEE Transactions on Knowledge and Data Engineering, 32*(2), 246–259.

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *Journal of Data and Information Quality (JDIQ), 11*(3), 1–37.

Zhu, W. Y., Peng, W. C., Chen, L. J., Zheng, K., & Zhou, X. (2016). Exploiting viral marketing for location promotion in location-based social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD), 11*(2), 1–28.