International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Foundations and Trends in Analytics and Big Data in Marketing

Tanmayee Tushar Parbat[1], Rohan Benhal[2], Honey Jain[1], Sajeeda Shikalgar[3]
[1]B.E IT, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India
[2]BBA IT, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India
[3]Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India

Article Info
Volume 8, Issue 5
Page Number: 100-107

Publication Issue :
September-October-2021

Article History
Accepted : 02 Oct 2021
Published : 26 Oct 2021

ABSTRACT

Given the research interest on Big Data in Marketing, we present a research literature analysis based on a text mining semi-automated approach with the goal of identifying the main trends in this domain. In particular, the analysis focuses on relevant terms and topics related with five dimensions: Big Data, Marketing, Geographic location of authors' affiliation (countries and continents), Products, and Sectors. A total of 1560 articles published from 2010 to 2015 were scrutinized. The findings revealed that research is bipartite between technological and research domains, with Big Data publications not clearly aligning cutting edge techniques toward Marketing benefits. Also, few inter-continental co-authored publications were found. Moreover, findings show that research in Big Data applications to Marketing is still in an embryonic stage, thus making it essential to develop more direct efforts toward business for Big Data to thrive in the Marketing arena.

Keywords : Big Data, Marketing, Geographic location

I. INTRODUCTION

Market dynamics, including its dialectic process of (de) regulation, (de) globalization, the last thirty-year information technology (IT) commoditization, the recent emergence of new revolutionary technology and national and international political uncertainties, beyond the cultural differences (Conti, Parente, & de Vasconcelos, 2015), alters the environmental dynamism and competitive advantage search. The advanced analysis with a marketing emphasis, denominated in the present work as marketing analytics, helps to transform internal or external data, structured or not, in strategic information. It demands some in-depth marketing modeling techniques for the market's response prediction, optimization of marketing-mix, and personalization for the customers (Wedel & Kannan, 2016). Data mining in texts, voice, video, digital media, or websites is a technology that helps organizations, providing insights that are used to adjust business rules and to create relationships with customers in a more relevant and connected way (Cooke & Zubcsek, 2017). Analytics, as a field of study, has been gaining momentum in the last two

decades, both in business and academic realms (Chen, Chiang & Storey, 2012). A search on Google Scholar in October 2019, with the keywords 'marketing' and 'analytics,' brought more than 470 thousand hits, with most of them consisting of recent publications. Some updated literature already predicts, for some industries, that emergent technologies and analytics will be enablers of competitive advantage and the organizations need to understand their data and prepare it for a more efficient use (Côrte-real, Oliveira, & Ruivo, 2017; Wang & Hajli, 2017; Braganza, Brooks, Nepelski, Ali, & Moro, 2017). The development of new products (Xu, Frankwick, & Ramirez, 2016), instant and recurrent feedback from transactions made by customers (Cooke & Zubcsek, 2017), and shared insights or co-creation innovation with customers (Khanagha, Volberda, & Oshri, 2017) are examples of activities supported by analytics technologies. There are also pricing and promotion, marketing mix, customer lifecycle value (Germann, Lilien, Fiedler, & Kraus, 2014), advertisement, salesforce, branding, positioning, and market segmentation, all guided toward data (Wedel & Kannan, 2016). From this, we can see that using analytics and analytical technologies for marketing has attracted much focus in research throughout the years, yet we can still notice a lack of full understanding of how the capabilities can associate analytics and marketing with performance. Therefore, the purpose of this study is to pave the path for further quantitative research in the field of analytics in marketing based on capabilities literature by suggesting the nomological network. For that, the paper addresses the following research questions: RQ1. How did the core knowledge at the intersection of marketing and analytics evolve through time, and what are the core topics and trends today? RQ2. What are the main research opportunities or issues when it comes to marketing capabilities? RQ3. What are the potential exogenous, covariates, intervenients, and concurrent endogenous constructs for studying the impact of adaptive analytics capabilities (proposed construct) on organizational performance? The present first presents the development of the analytics research, with the focus on its use in marketing; then, it is presented some descriptive bibliometric results, after two cluster analysis for author coupling and keyword co-occurrence; then it is established a systematic review of recent quantitative papers from where, finally, it is proposed a nomological network that shows the pathways to improve quantitative research in marketing strategy using analytics.

II. MARKETING AND ANALYTICS DUO: FOUNDATIONS

The present work provides a twenty-year summary of the analytics development as a research field while highlighting the major strengths of combining this area with marketing, especially from the perspective of the capabilities literature. The broad term analytics is considered a young but increasingly important field of study, mainly characterized as the set of techniques, tools, and approaches aiming at accurate analysis of business data to improve decision-making. The evolution of this multidisciplinary field can be simplified into three major time periods, as proposed by Chen et al. (2012) framework: BI&A 1.0, BI&A 2.0, and BI&A 3.0. The first period occurs in the 1990s, where the combination of statistical techniques and data mining practices lead to the development of better analytical tools, designed for the extraction and storage of data into robust databases (Chen et al., 2012). Therefore, the goal of marketing professionals was to optimize the collection, structuring, and later analysis of the data available in the market.

The third and most recent period is directly related to the so-called internet of things (IoT) and big data phenomena. This opens a new perspective on how new high-tech products can be used as a data source to provide useful and individualized information to enhance market knowledge, though it also brings uncertainty about the best techniques and approaches

to collect, process, and analyze data (Chen et al., 2012).

III. LITERATURE REVIEW

It was performed a bibliometric study in Scopus and Web of Science bases to analyze the state of the art that relates marketing and analytics. The sample criteria were created using "marketing AND analytics AND (capabilities OR resources)"as the search string in titles, abstracts, and keywords. English only, and with no period of time limit criteria. The keywords capabilities or resources were chosen for the theoretical delimitation of the present work, which used the resource-based view and its underlying literature about capabilities as cornerstones. It was confirmed that the extracted articles are adherent to the research delimitation with a search based on the titles, abstracts, and keywords. Those were passed onto a spreadsheet, from where other theories were found with low occurrences, such as the configuration theory, game theory, innovations theory, while, with a bigger number of occurrences, it was perceived capabilities, Resource-Based Theory, or Resource-Based View (RBV). According to Quevedo-Silva, Santos, Brandão, and Vils (2016), the bibliometric method brings a broader comprehension of themes or areas, allowing the identification of trends. According to the bibliographic review of those authors, this type of study performs three, mutually non-exclusive, approaches: (i) descriptive, that draws broad lines or the most studied topics in an area, identifying research groups, publication year, leading authors and methods used; (ii) about methodology, which ought to understand the methodological domain of the researched area by classifying and counting the research drawings and the test techniques mostly used, adjacently, it aims to find study opportunities and highlight a determined area research tradition; lastly the approach of (iii) descriptive analysis, which deepens the knowledge using cluster analysis of authors, theories, and keywords from the sample metadata. The descriptive step was used to introduce the present bibliometric study, but the second step was not performed with a bibliometric approach, due to the research delimitation. Regarding the methodological domain, it was not carried out on all 705 articles, because it was chosen to perform this analysis in a much more detailed manner, beginning with a shorter list presented in the applied research section. Finally, the last step, cluster analysis, is shown as research trends.

IV. METHODOLOGY

Because most of the papers are theoretical or do not use frequent management quantitative methods, it was also performed an applied bibliographical revision in international journals with significant JCR indexes, deliberated as more than one, and available in periodicos.capes.gov.br. The searched constructs were related to “Performance Management,” “Technology and Learning,” or “Customer Orientation.” For this applied bibliographical revision, we followed Hong, Chan, Thong, Chasalow, and Dhillon (2013) guidelines, about contextual constructs, i.e., similar constructs with different names depending on context, the present work considered terms like big data analytics, social media analytics, marketing analytics, and customer analytics as constructs related to specific contexts from the general construct business analytics, following the Chen et al. (2012) approach. It was necessary to analyze quantitative works using all these terms. By delimitation, it was included only the papers that exhibited the structural model and their respective measurements. Other works, not initially selected, were included by ramification, like a snowball (similar to convenience sampling) from the articles initially chosen. The result is presented in Table 8, which is the summary of this applied bibliographic review with pertinent and explicit quantitative models only. It shows the constructs and a brief description of the general idea of each work. Table 8 reveals a list of selected applied research with explicit

quantitative models since the year 2010, from where future studies can be operationalized using the nomological network constructs options (see Figure 4). Marketing analytics can be studied from the capabilities literature point of view (Germann et al., 2014; Wamba et al., 2017), but a literature mapping shows different levels of abstractions and some challenging issues (see Figure 1). Figure 1 helps to explain the present work's theoretical background/choices.

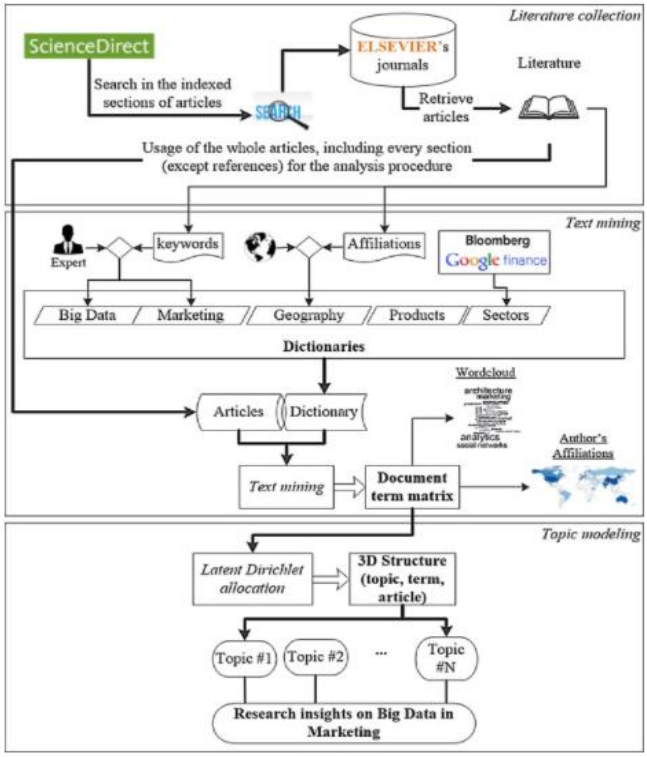

Fig. 1: Schematic of the proposed text mining and topic modeling approach

Stemming is a technique often applied in Text Mining, in order to reduce similar words to a unique term (e.g., "predicting" and "predictive" are transformed in "predict"). Rather than just performing usual stemming, an extended list of related terms was created that includes other concepts in the same domain.

The definition of dictionaries and the grouping of terms under a unique reduced term are subjective. To try reducing this subjectivity, several enhancements were introduced to the original method. The most significant is the usage of articles keywords to build both the Big Data and Marketing dictionaries, as shown in Fig. 1. Then the result of merging together all keywords for building the final dictionary was assessed. Examples of such keywords are "big data" and "massive data".

The geographical dictionary was defined through the association of countries and cities to continents. This association was made with the author's information, extracting their address from each article with a text mining algorithm. As an example, if an article has authors from "Portugal" and "Brazil" then it belongs to two continents: "Europe" and "South America".

The sectors' dictionary allows to identify the main sectors of the economy considered in the articles set. This dictionary was defined by consulting information on how the sectors are divided on Bloomberg and Google Finance.

The products' dictionary intends to identify the principal products, divided into physical goods and intangible services, which are addressed in articles. This dictionary was defined with the help of the sectors dictionary, and also manually including related words. Each dictionary is a major input to the text mining procedure; the other is the full content of the articles (with exception of the references sections), as shown in Fig. 1.

V. DATA ANALYSIS

Table 1 presents the cross-domain topics for Big Data and Marketing, in a total of eighteen. The results show an increasing interest in Big Data applied to Marketing. For each topic there is always a dominant term, with a value that matches it closer to a certain Marketing question or to a type of Big Data technique, tool or context. A deeper analysis from the three most relevant terms per topic provides an interesting insight: there are a large number of topics mainly characterized by three terms all related with Big Data (four topics, numbers 4, 9, 16 and 18, with a total of 236 articles) and with all three terms associated with

Marketing (eight topics, numbers 2, 5, 6, 7, 8, 12, 14 and 17, with a total of 701 articles). Such a result means that from the total of 1560 articles collected, 937 (around 60%) are mostly focused on just one of the domains, limiting cross-disciplinary research. This is an interesting finding, revealing a large gap for Big Data research that demonstrates the benefit to the Marketing discipline. Particularly relevant is the fact that Big Data emerges quite often in Marketing related publications, although there seems to exist low emphasis on the technical aspects that help to improve Marketing issues.

Table 1: Relevant topics for Big Data in Marketing

T	Nr.	Term 1	$\beta 1$	Term 2	$\beta 2$	Term 3	$\beta 3$	2010	2011	2012	201
1	260	Algorithm	0.00	Data processing	7.08	Customer segmentation	7.80	7	13	18	33
2	175	Market	0.02	Sales	4.40	Price	5.05	5	10	11	20
3	155	Architecture	0.02	Data processing	3.98	Lifetime value	8.29	12	7	9	18
4	139	Big Data	0.15	Data processing	2.32	Nosql	3.35	0	0	2	6
5	126	Media	0.01	Marketing	4.55	Advertising	6.53	2	3	6	17
6	119	Product	0.07	Product development	3.42	Market	3.99	3	6	6	13
7	94	Marketing	0.55	Sales	1.06	B2b	3.19	2	4	7	9
8	78	Price	0.03	Market	3.70	Sales	5.48	5	4	7	11
9	67	Visualization	0.96	Mapreduce	1.12	Hadoop	1.29	3	1	2	3
10	65	Modeling	0.01	Visualization	5.42	Market	5.48	3	3	4	8

From the remaining six cross-disciplinary topics (accounting for the remaining 40% of articles), the first, with the largest number of articles, is mainly focused on Big Data algorithms, although it conceals a weak relationship with both "data processing" and "customer segmentation", showing a relevant specific Marketing task such as segmentation which is highly dependent on data. A similar result occurs for topic number three, where architecture is weakly related to "lifetime value", an apparently awkward connection that may be justified by the second term, "data processing", which is a key asset toward an accurate evaluation of customer lifetime value. Topics number 10 and 11 are characterized by two of the most widely used broader terms when it comes to extracting value from data: "modeling" and "prediction". Also, both highlight the usage of these terms to "market". However, the latter holds a moderate relationship with "customer retention", which is a specialized Marketing task. Likewise, topic number 13 also conceals a relation from "analytics" with "market". Thus, three inter-disciplinary topics encompassing 182 articles (nearly 12% of the total) have a relationship with market, disregarding other more specific Marketing tasks. Topic number fifteen discovered a relationship between "social networks" and "data processing"; as the volumes of data keep increasing exponentially, it is crucial to quickly handle that data for extracting useful knowledge. Generally, the findings show evidence of the lack of real inter-disciplinary research that equally addresses problems and methods from both domains, calling for more applied research aiming to understand how Big Data is changing established paradigms in Marketing.

Table 2 provides valuable insights on authors' affiliations regarding geographic location. It clarifies if authors are producing worldwide scale cross research on Big Data in Marketing, regarding distribution of continents. The results show that both North American and European researchers are not developing research with co-authors from other continents, given that both topics 1 and 3 achieve a β value of zero for Europe and North America respectively, representing a perfect match to these continents. The remaining two topics are closer related to Asia and Oceania, while also fitting articles published by South American and African authors, respectively.

Table 2: Relevant topics for Geographical locations

T	Nr.	Term 1	β1	Term 2	β2	Term 3	β3	2010	2011	2012	2013	2014	20
1	612	Europe	0.00	South America	23.91	North America	75.19	24	21	38	73	149	30
2	497	Asia	0.03	South America	3.41	North America	61.63	10	14	23	40	117	29
3	364	North America	0.00	Europe	46.34	South America	48.80	12	20	27	45	106	15
4	87	Oceania	0.22	Africa	1.82	Central America and Caribbean	3.30	2	6	4	10	21	44

VI. CONCLUSION

The findings from this study rise in the form of prescriptions for future research. First, while plenty of research is being conducted on Big Data and on Marketing, less is found in addressing specifically the benefits that marketers could potentially achieve through Big Data solutions. While Big Data adoption within the industry is taking place nowadays, there is a gap for research to clearly identify the pros and cons for organizations to invest in Big Data. After defining the boundaries for a Big Data solution, it is imperative that the implementation is perfectly aligned with the challenges posed by the specificities of the business, as each solution needs to be context-aware. The confirmation of such discovery unveiled a research gap in cross-disciplinary research, with technological researchers needing to better align the benefits of Big Data toward Marketing. It is interesting to note that although several specific Marketing related terms often seen as associated to data analysis were included (e.g., customer retention, customer segmentation), few appear highlighted in the topics uncovered, and the ones that did appear, show a weak relationship to the corresponding topic, paling in comparison to the relevance of the respective Big Data term.

VII. REFERENCES

[1] Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007.

[2] Barney, J. B. (2014). How marketing scholars might help address issues in resource-based theory. Journal of the Academy of Marketing Science, 42(1), 24–26. https://doi.org/10.1007/s11747-013-0351-8.

[3] Barrales-Molina, V., Martínez-López, F. J., & Gázquez-Abad, J. C. (2014). Dynamic marketing capabilities: Toward an integrative framework. International Journal of Management Reviews, 16(4), 397–416. https://doi.org/10.1111/ijmr.12026.

[4] Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. Journal of Business Research, 70, 328–337. https://doi.org/10.1016/j.jbusres.2016.08.006.

[5] Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. Social Science Information, 22(2), 191–235. https://doi.org/10.1177/053901883022002003.

[6] Chandler, J. D., & Wieland, H. (2010). Embedded Relationships: Implications for Networks, Innovation, and Ecosystems. Journal of Business Market Management, 4(4), 199–215. https://doi.org/10.1007/s12087-010-0041-5.

[7] Chang, W., Park, J. E., & Chaiy, S. (2010). How does CRM technology transform into organizational performance? A mediating role of marketing capability. Journal of Business Research, 63(8), 849–855. https://doi.org/10.1016/j.jbusres.2009.07.003.

[8] Chen, L., & Nath, R. (2018). Business analytics maturity of firms: an examination of the relationships between managerial perception of IT, business analytics maturity and success. Information Systems Management, 35(1), 62–77.

[9] Johnson, J. S., Friend, S. B., & Lee, H. S. (2017). Big Data Facilitation, Utilization, and Monetization: Exploring the 3Vs in a New Product Development Process. Journal of Product Innovation Management, 34(5), 640–658. https://doi.org/10.1111/jpim.12397.

[10] Kessler, M. M. (1963). Bibliographic Coupling Between Scientific Papers. American Documentation, 14(1), 10–25. Khanagha, S., Volberda, H., & Oshri, I. (2017). Customer Co-Creation and Exploration of Emerging Technologies: The Mediating Role of Managerial Attention and Initiatives. Long Range Planning, 50, 221–242. https://doi.org/10.1016/j.lrp.2015.12.019.

[11] Kim, H. J., Jo, N. O., & Shin, K. S. (2016).Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. Expert Systems with Applications, 59, 226–234. https://doi.org/10.1016/j.eswa.2016.04.027.

[12] Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. Journal of Marketing, 80(1), 7–25. https://doi.org/10.1509/jm.14.0249.

[13] MacKenzie, Podsakoff, & Podsakoff. (2011). Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. MIS Quarterly, 35(2), 293. https://doi.org/10.2307/23044045.

[14] Rosati, P., Cummins, M., Deeney, P., Gogolin, F., van der Werff, L., & Lynn, T. (2017). The effect of data breach announcements beyond the stock price: Empirical evidence on market activity. International Review of Financial Analysis, 49, 146–154. https://doi.org/10.1016/j.irfa.2017.01.001.

[15] Sheng, J., Amankwah-amoah, J., & Wang, X. (2017).A multidisciplinary perspective of big data in management research. International Journal of Production Economics, 191(November 2016), 97–112. https://doi.org/10.1016/j.ijpe.2017.06.006.

[16] Tan, K. H., & Zhan, Y. (2017). Improving new product development using big data: a case study of an electronics company. R and D Management, 47(4), 570–582. https://doi.org/10.1111/radm.12242.

[17] Tan, K. H., Zhan, Y. Z., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. International Journal of Production Economics, 165(July 2015), 223–233. https://doi.org/10.1016/j.ijpe.2014.12.034.

[18] Teo, T. S. H., Nishant, R., & Koh, P. B. L. (2016). Do shareholders favor business analytics announcements? Journal of Strategic Information Systems, 25(4), 259–276. https://doi.org/10.1016/j.jsis.2016.05.001.

[19] Trainor, K. J., Andzulis, J., Rapp, A., & Agnihotri, R. (2014). Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. Journal of Business Research, 67(6), 1201–1208. https://doi.org/10.1016/j.jbusres.2013.05.002.

[20] Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 70, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009.

[21] Wang, Y., & Byrd, T. A. (2017).Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. Journal of Knowledge Management, 21(3), 517–539. https://doi.org/10.1108/JKM-08-2015-0301.

[22] Wang, Y., & Hajli, N. (2017).Exploring the path to big data analytics success in healthcare. Journal of Business Research, 70, 287–299. https://doi.org/10.1016/j.jbusres.2016.08.002.

Cite this article as :

Tanmayee Tushar Parbat, Rohan Benhal, Honey Jain, Sajeeda Shikalgar, "Foundations and Trends in Analytics and Big Data in Marketing", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 5, pp. 100-107, September-October 2021.
Journal URL : https://ijsrcseit.com/CSEIT217681