2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA)

SiloSolver: Developing an Algorithm for Automatic Aggregation and Testing of Isolated Customer Segments in Facebook Ads Campaigns

Joni Salminen
Qatar Computing Research Institute
Doha, Qatar
jsalminen@hbku.edu.qa

Tommi Salenius
Independent researcher
Turku, Finland
tommi@tommisalenius.com

Bernard J. Jansen
Qatar Computing Research Institute
Doha, Qatar
jjansen@acm.org

***Abstract*— Data silo problem refers to related datasets located in different databases, systems, or files. In the case of online advertising, performance metrics are fragmented in different campaigns and ad sets, making it difficult to compare customer segments. In this research, we present *SiloSolver*, an algorithm that (a) retrieves performance metrics for different customer segments across all campaigns, (b) aggregates the values for each customer segment in mean values, and (c) runs a statistical comparison (Student's t-test) on the performance differences between the segments. The algorithm is evaluated using a real Facebook Ads dataset from an e-commerce company consisting of hundreds of campaigns from over five years. Using *SiloSolver*, advertisers using Facebook Ads are better able to understand their market segments across multiple seemingly disparate campaigns.**

Keywords Facebook Ads, online advertising, digital marketing, customer segmentation

I. Introduction

Marketers desire to know the best customer segments to target with their ads [1]–[3]. However, the information about the performance of different customer segments [4] can easily get lost across many campaigns, ad platforms, and time points [5]. This problem is particularly prominent in online advertising, where marketers tend to (a) use many different platforms, such as Google Ads, Facebook Ads, Twitter Ads, etc., and (b) create many campaigns with different creatives, target groups, and experimental conditions.

This conundrum results in a scenario where the information about the same customer segment (e.g., Women, 25-34, Tallinn) is located in different places, repositories, and databases, making it difficult to get an understanding of a segment's cross-campaign marketing performance. This silo problem is particularly vexing in the case of online advertising platforms, such as Facebook Ads, where the performance metrics pertaining to various customer segments are fragmented across dozens or hundreds of campaigns and ad sets. This fragmentation is common even for small-medium size businesses (SMBs) for at least two reasons: because (a) Facebook's algorithm can automatically target hundreds of customer segments, and (b) advertisers often create new "silos" (i.e., campaigns, ad sets) because this is easy to do in the platform and due to nature of their ongoing promotional campaigns.

Over time, these additional campaigns accumulate into a large number of isolated information sets about different customer segments' behavior. Consider, for example, that Segment 1 would be presented with 1000 ad impressions in *Campaign A* and 1000 ad impressions in *Campaign B*. In Facebook Ads' user interface (UI), it is easy to analyze both campaigns separately, but it is not possible to easily combine their information in real-time for the analysis. Clearly, doing so would be useful because the marketer could double the statistical power of their analyses (in the above example: 1000 + 1000 = 2000 impressions).

In the absence of automated solutions, of which we currently know of none, the segment data must be retrieved, exported, and analyzed manually. The more campaigns there are, the more manual work is needed for data integration and analysis. Considering that Facebook Ads is, at the time of the study, used by more than five million advertisers [6], the manual work in combining the segments' data results in a large number of work hours in the online ads industry. In addition, it is commonly stated that data preparation activities tend to take up to 80% of work time in analysis projects [7], so it is important to try and reduce this workload to free marketers' time for more strategic activities.

As a potential solution for this isolated customer segments problem, we propose an algorithm dubbed *SiloSolver*. This algorithm performs a simple but impactful functionality: it (a) retrieves various performance metrics for different customer segments across all campaigns, (b) aggregates the values for each customer segment in mean values, and (c) performs a statistical comparison (Student's t-test) to measure the differences between the mean values of the segments.

By applying *SiloSolver*, online advertisers using Facebook Ads can answer important questions such as:

- *Across all campaigns, which target groups perform the best?*
- *How are all campaigns perform for target groups?*

To demonstrate the applicability of *SiloSolver* across different marketing goals, we examine three groups of advertising goals with their associated metrics: (a) *awareness* (metric: reach), (b) *engagement* (metric: clicks), and (c) *conversion* (metric: purchases). These goals are typical and important marketing goals [8].

We evaluate the approach using a real Facebook Ads dataset from a Finnish e-commerce company, consisting of hundreds of campaigns from over five years' period. Furthermore, we publish the *R* source code for *SiloSolver* for further research and development (https://github.com/joolsa/silosolver) .

The main implication of this research is that with *SiloSolver*, online marketers using Facebook Ads can better understand the overall performance of their campaign relative to different target groups and make adjustments to achieve business objectives from this understanding. We also provide empirical insights from the dataset, observing that the most lucrative target groups frequently change when changing the performance metric. This suggests that choosing the optimal

978-1-6654-2180-5/21/$31.00 ©2021 IEEE

targeting criteria depends on the marketing objective; i.e., different segments are optimal for different goals. Moreover, *SiloSolver* exemplifies the use of automation for marketing tasks – while the algorithm in itself includes no notable "artificial intelligence" (AI) [9], it does compute statistical functions similar to those in machine learning (ML) and could, with further development, be deployed as a part of an intelligent advertising system [10].

II. Related Literature

One of the earliest advocates of customer segmentation is Smith [3], who, in 1956, emphasized the importance of market segmentation along with product segmentation. Since its conception, customer segmentation has been an ongoing research area [1], [11]–[13]. Conceptually, customer segments involve attributes that unify groups of customers or separate these groups from others [14]. The segments can be described by their core demographic or other properties [15], or they can be personified and enriched into humanlike segment representations for customer-centric decision making [16].

The availability of online customer data has greatly increased in digital environments [17], providing potential for innovative customer segmentation solutions [18], such as creating data-driven personas [19]. In turn, the high volume of customer data has increased the difficulty of turning this data into knowledge and insights that serve online marketers and other business stakeholders [20]–[22]. This issue of "too much data in too many places" has particularly been recognized in online marketing, where decision makers struggle to make sense of fragmented data on customer behavior [5] and to identify the segments that matter more than others. Thus, even though there has been a variety of algorithms and methods to create customer segments [2], [23], [24], novel solutions to deal with fragmented customer data in ways that facilitate the knowledge work of online advertisers are required.

Moreover, despite being the second largest online advertising platform after Google Ads [25], there is surprisingly little academic research focusing on Facebook advertising as a tool of customer segmentation. Nevertheless, there is a tremendous need for such research, considering that Facebook Ads is used by more than five million active advertisers [6] across the globe, and the platform operates advertising budgets totaling to more than 17 billion dollars at the time of the study [25]. Therefore, any research that facilitates the jobs or advertisers using Facebook Ads has the opportunity to provide economies of scale and scope in the marketing profession.

This research is among the first practice-oriented studies to describe the functionalities of Facebook advertising and particularly address a key shortcoming of the platform from the advertisers' point of view. Thus far, academic reports of applying automation for marketing have been limited, even though the potential of automation in marketing was discussed already in the 1960s [26], [27]. Despite the revitalized interest in AI [9] and ML [28] in the marketing community, academic studies reporting different tools for facilitating the daily jobs of online advertisers remain limited. To this end, our study provides an example of how automation can be used to facilitate the analysis of customer segments in online advertising.

III. Methodology

A. Intuition of the Algorithm

SiloSolver addresses the data siloing problem in Facebook Ads by retrieving and aggregating performance information on the customer segments observed in the data.

The algorithm analyzes all observed target group combinations (*age* x *gender* x *location*), groups the data, sorts based on performance, and runs statistical calculations (1-tailed t-test) to determine if the differences are statistically valid and not due to chance. Overall, the *SiloSolver* algorithm relies on six main steps:

1. **STEP 1:** creating unique target group combinations retrieved from Facebook Ads platform
2. **STEP 2:** retrieving the unique campaigns from the data, along with the performance metrics by each segment in each campaign
3. **STEP 3:** aggregating the mean values of the performance metrics from each campaign by customer segments
4. **STEP 4:** sorting each customer segment from highest to lowest mean value by each performance metric
5. **STEP 5:** statistically testing the N highest-performing groups against other (N-1) groups using t-tests, and
6. **STEP 6:** printing the results in an output file

The following sections provide details of the process.

B. Defining Customer Segments

Defining the customer segment, we focus on three demographic criteria available in the Facebook platform (and in most other online ad platforms): *age*, *gender*, and *location*. Age is available at a group level, gender includes three possible levels (male/female/unknown), and locations are the regions within the target market (e.g., counties in Finland). To calculate the maximum number of segments, we apply the following formula:

$$\textbf{Segments} = \text{age}_n \times \text{gender}_n \times \text{location}_n$$

For example, in the Finnish e-commerce dataset that we utilize for the evaluation (see Section 8), there are eight age groups (13-17; 18-24; 25-34; 35-44; 45-54; 55-64; 65+; Unknown), 3 genders (male, female, unknown—these three genders are available on Facebook Ads and do not reflect the researchers' views on gender politics), and 957 locations, yielding 8 x 3 x 957 = 22,968 possible combinations. The real number of combinations is lower, however, because not all combinations have observations in the data. Note that the algorithm is generalizable to other Facebook Ads datasets because the number of unique levels of the three targeting criteria can be trivially obtained from each dataset at hand.

C. Retrieving Campaign Data

Data can be obtained in two ways from Facebook Ads: (a) by manually exporting comma-separated values (CSV) files from the Facebook Ads Manager (see Figure 1), or (b) programmatically via the Facebook Marketing API (https://developers.facebook.com/docs/marketing-apis/). To evaluate the algorithm, we obtained the data by manually exporting it from Facebook Ads Manager.

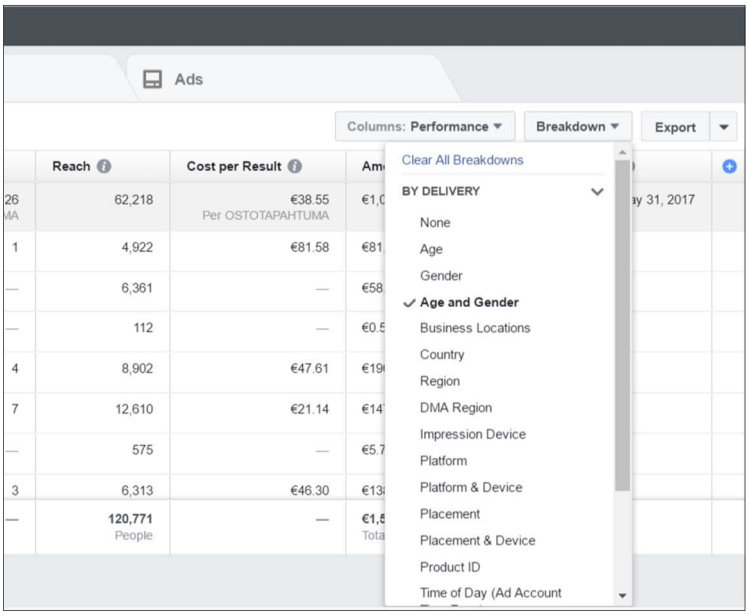

Fig. 1. Facebook Ads export interface.

D. Selecting Performance Metrics

Facebook Ads provides dozens of performance metrics. Analyzing all of them is not crucial or relevant for an e-commerce company. Therefore, we retrieve only metrics relevant for a typical e-commerce marketing scenario. The typical e-commerce marketing scenario involves campaign objectives relating to generating awareness, interest, and action from the customers [29]. Table 1 explains the metrics corresponding to these objectives.

TABLE I. EXPLANATIONS OF THE VARIABLES

Variable	Explanation
Cost-per-click (CPC)	Cost in currency (dollar, euro…) of a click by a user on an ad. Calculated by dividing the cost of the campaign with the number of clicks.
Click-through rate (CTR)	The number of people who clicked the ad after having seen it.
Conversion rate (CVR)	The number of people who purchased out of the ones who clicked the ad's website link.

Based on the chosen metrics, the *SiloSolver* computes three different customer segment groups:

- **Cheapest to reach (CPC)** = lowest mean value for avg. CPC
- **Most likely to click (CTR)** = highest mean value for avg. CTR
- **Most likely to buy (CVR)** = highest mean value for avg. CVR

These groups were chosen because they reflect the objectives of online advertising—e.g., reach corresponds to brand awareness, clicks correspond to response or interest, and purchases correspond to the sales objective [5], [8]. The UIs of online ad platforms, including that of Facebook Ads, are increasingly directing advertisers to choose their goals prior to any other decisions; hence, goal-based segmentation conforms to this paradigm of doing digital marketing.

We choose relative metrics to account for exposure bias [30]. In other words, a given group can show higher performance in absolute numbers (e.g., clicks) because it is more frequently targeted, but it can perform relatively worse when compared to another group by CTR.

E. Statistical Testing

The purpose of the testing is to evaluate the reliability of the differences between the customer segments in the data. We use a one-tailed Student's t-test that provides more statistical power than a two-tailed test. One-tailed testing is appropriate since we are interested if a segment performs *better* than another segment. We test the following hypothesis:

H1a: *The mean value of the performance metric tested is greater for the segment tested (Group 1) than for the other segments (Group 2).*

In other words, the null hypothesis is: **H0:** *The mean values of the performance metric between the two segments are equal.*

An exception is the CPC metric, which uses 'less' for t-test comparison, because lower CPC is more desirable for the advertiser, all else being equal. Therefore, the formulation for the CPC metric is: **H1b:** *The mean value of the performance metric tested is lower for the segment tested (Group 1) than for the other segments (Group 2).*

We compare two groups: the first group is the highest performing customer segment for a given performance metric. The second group, against which we compare, consists of mean values of all the other groups in the top-*N* set of highest performing segments. The top-N set is obtained by aggregating the performance data across the different campaigns for each demographic group and then sorting from highest to lowest.

In other words, we compare if the mean of a performance metric for a given customer segment belonging to the top-*N* set is greater than the aggregated mean of all the other segments belonging to the top-*N* set.

In *SiloSolver*, N is a parameter that can be adjusted—however, we suggest a relatively low value, such as $N = 5$, to limit computational complexity and to avoid the manageability and information overload problems associated with showing too many segments to end users [31]. The algorithm iterates across each customer segment belonging to the top-N set, as well as iterating for each of the set performance metrics (see Table 1). In other words, we repeat the t-test with each segment belonging to the top-N set, where the other segment always consists of the aggregated mean of the other segments—i.e., excluding the group whose performance we are testing against. We filter out groups with less than ten observations [32], which is the minimum recommended sample size for statistical calculations.

Note that the groups themselves contain several individual people; one group observation in the data can include several thousands of ad impressions, but it is only calculated as one observation in *SiloSolver*, because Facebook does not provide access to individual level data. In this sense, the data we utilize has "hidden" statistical power that cannot be directly accessed due to limits of aggregation in the ad platform, but that still can (theoretically) improve the robustness of the results.

IV. Evaluation on a Real Dataset

We evaluate the *SiloSolver* algorithm with a real dataset collected from Facebook Ads of a Finnish e-commerce company that is selling experience gifts and is actively using Facebook advertising to generate sales revenue. The dataset contains 620 individual campaigns carried out over the span of five years. The campaign objectives are varied, including sales conversions, brand awareness, engagement, and so on.

Due to the large number of campaigns and objectives, this dataset was appropriate for demonstrating the usefulness of our approach. Without automation, a marketer would need to manually choose the 620 campaigns and separately export data from each, then combine it by copy-pasting, and finally conduct statistical analyses one by one.

As explained in Section 7, *SiloSolver* produces Top N segments for each metric. Here, we set the $N = 5$, as (a) this number is enough to demonstrate the capabilities of the algorithm, (b) retains computational complexity manageable (i.e., makes the algorithm run in a reasonable time), and (c) addresses the cognitive capacity issue regarding too many segments for efficient decision-making [11].

Table 2 displays the results from applying *SiloSolver* to the evaluation dataset. From Table 2, we can observe that the best-performing segments are different for each performance metric. This suggests that the optimal customer segments vary greatly by the chosen performance metric. However, both hypotheses are supported in that we can observe differences between two groups and differences among all segments for a chosen campaign performance metric.

V. Implications and Future Work

There is a need for computational methods that aggregate the fragmented data on customer segments. Moreover, there is a need for statistical methods to establish the reliability of the observed performance differences among the customer segments. In this research, we have presented *SiloSolver*, an algorithmic approach that addresses both problems.

TABLE II. Results of *SiloSolver* from the evaluation dataset.

Cheapest to reach (CPC)									
Age	***Gender***	***Location***	***N***	***Avg.Amount.Spent***	***Avg.Link.Clicks***	***CPC***	***Rank***	***t.test***	***p.value***
13-17	male	Northern Ostrobothnia	28	0.005	0.200	0.023	1	-0.362	0.376
13-17	female	Northern Ostrobothnia	30	0.007	0.168	0.045	2	0.858	0.760
13-17	male	Southwest Finland	32	0.147	0.522	0.281	3	-0.383	0.369
18-24	male	Päijät-Häme	11	0.648	2.264	0.286	4	-3.347	**0.003**
18-24	male	South Karelia	11	0.183	0.630	0.290	5	-5.674	**0.001**
Most likely to click (CTR)									
Age	***Gender***	***Location***	***N***	***Avg.Link.Clicks***	***Avg.Impressions***	***CTR***	***Rank***	***t.test***	***p.value***
45-54	female	Päijät-Häme	13	11.619	872.505	0.013	1	2.029	**0.032**
45-54	female	Satakunta	13	9.830	756.862	0.013	2	1.255	0.117
18-24	male	Kymenlaakso	11	1.591	126.303	0.013	3	2.581	**0.012**
45-54	female	Kymenlaakso	13	8.386	668.730	0.013	4	1.250	0.117
35-44	male	Tavastia Proper	12	9.958	806.034	0.012	5	2.380	**0.016**
Most likely to buy (CVR)									
Age	***Gender***	***Location***	***N***	***Avg.Website.Purchases***	***Avg.Link.Clicks***	***CVR***	***Rank***	***t.test***	***p.value***
45-54	female	Åland Islands	87	0.008	0.020	0.408	1	0.957	0.183
35-44	female	Åland Islands	111	0.006	0.027	0.225	2	0.961	0.181
55-64	female	Åland Islands	69	0.002	0.012	0.168	3	0.990	0.180
55-64	male	Åland Islands	65	0.001	0.008	0.091	4	0.998	0.182
18-24	female	Etelä-Savo	11	0.076	1.100	0.069	5	1.278	0.119

We used an evaluation dataset of 620 real campaigns spanning five years of active advertising in Facebook Ads and found that the highest performing segments vary by performance metric. *SiloSolver* automates the analysis that would a long time if done manually, and even much more when analyzing Facebook Ads accounts with thousands of campaigns.

With *SiloSolver*, online marketers using Facebook Ads can better understand the performance of their customer segments. Moreover, the algorithm can help avoid biased interpretations in the cases where data on smaller segments are scattered across the campaigns. By grouping the disparate data, marketers achieve a higher power for statistical analysis than with using isolated data. By not focusing only on the largest customer segments, online marketers can reach a more comprehensive understanding of their diverse customer segments.

SiloSolver provides a list of the optimal segments for a given campaign objective. These segments can then be targeted in future advertising, both in Facebook and beyond. Based on the chosen customer understanding goal, different metrics can be deployed to segment the same customer base in alternative ways; for example:

- "**Cheapest to reach**" => based on the lowest CPC
- "**Most likely to click**" => based on the highest CTR (calculated metric)
- "**Most likely to buy**" => based on the highest CVR (calculated metric)
- "**Highest sales value**" => based on the highest Total conversion value (EUR)
- "**Biggest engagement group**" => based on the highest Clicks (All)
- "**Biggest visitor group**" => based on the highest number of Link clicks
- "**Most buying**" => based on the highest number of Website purchases

For example, the experience gift e-commerce company applied the algorithm in two ways: to (a) *validate* their existing beliefs about the most profitable target groups (young women in the Uusimaa region in Finland), and (b) *diversify* their marketing efforts to other groups for brand awareness campaigns, because the "cheapest to reach" segments are different from the most profitable segments.

Future development of *SiloSolver* could include programmatic access to Facebook Marketing API or Facebook Audience Insights (see Fig. 2) in order to automate the data retrieval process. The automation of *SiloSolver* is limited to automatically retrieving the values of the performance metrics for each target group and aggregating them for statistical testing.

The key limitation of *SiloSolver* relates to data quality of Facebook Ads. As noted in earlier research [11], this data is aggregated, and marketers are unable to validate its correctness for individual consumers [33]. Another limitation is that the current implementation of the algorithm is applicable only to the Facebook Ads platform, not e.g., Google Ads.

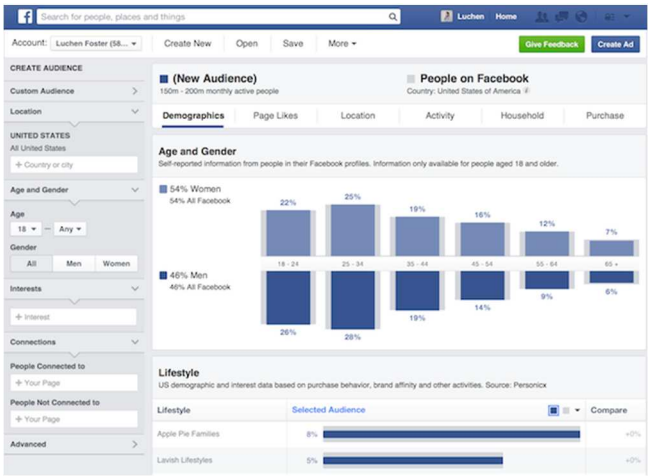

Fig. 2. Screenshot of Facebook Audience Insights.

REFERENCES

[1] B. P. Shapiro and T. V. Bonoma, "How to Segment Industrial Markets," *Harvard Business Review*, May 01, 1984. https://hbr.org/1984/05/how-to-segment-industrial-markets (accessed Dec. 03, 2017).

[2] T. Jiang and A. Tuzhilin, "Segmenting Customers from Population to Individuals: Does 1-to-1 Keep Your Customers Forever?," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 10, pp. 1297–1311, Oct. 2006, doi: 10.1109/TKDE.2006.164.

[3] W. R. Smith, "Product Differentiation and Market Segmentation as Alternative Marketing Strategies," *Journal of Marketing*, vol. 21, no. 1, pp. 3–8, 1956, doi: 10.2307/1247695.

[4] B. J. Jansen, S. Jung, D. Ramirez Robillos, and J. Salminen, "Too Few, Too Many, Just Right: Creating the Necessary Number of Segments for Large Online Customer Populations," *Electronic Commerce Research and Applications*, p. 101083, Jul. 2021, doi: 10.1016/j.elerap.2021.101083.

[5] J. Järvinen and H. Karjaluoto, "The use of Web analytics for digital marketing performance measurement," *Industrial Marketing Management*, vol. 50, no. Supplement C, pp. 117–127, Oct. 2015, doi: 10.1016/j.indmarman.2015.04.009.

[6] Business Insider, "Facebook adds a million advertisers in 7 months," *Business Insider*, Apr. 12, 2017. https://www.businessinsider.com/facebook-adds-a-million-advertisers-in-7-months-2017-4 (accessed Oct. 04, 2018).

[7] P. N. Mahalle, G. R. Shinde, P. D. Pise, and J. Y. Deshmukh, "Data Collection and Preparation," in *Foundations of Data Science for Engineering Problem Solving*, Springer, 2022, pp. 15–31.

[8] T. B. Clarke and B. J. Jansen, "Conversion potential: a metric for evaluating search engine advertising performance," *Journal of Research in Interactive Marketing*, vol. 11, no. 2, pp. 142–159, 2017.

[9] M. Mustak, J. Salminen, L. Plé, and J. Wirtz, "Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda," *Journal of Business Research*, vol. 124, pp. 389–404, Jan. 2021, doi: 10.1016/j.jbusres.2020.10.044.

[10] Y. Yang, Y. C. Yang, B. J. Jansen, and M. Lalmas, "Computational Advertising: A Paradigm Shift for Advertising and Marketing?," *IEEE Intelligent Systems*, vol. 32, no. 3, pp. 3–6, May 2017, doi: 10.1109/MIS.2017.58.

[11] J. An, H. Kwak, S. Jung, J. Salminen, and B. J. Jansen, "Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data," *Social Network Analysis and Mining*, vol. 8, no. 1, 2018, doi: 10.1007/s13278-018-0531-0.

[12] P. R. Dickson, "Person-Situation: Segmentation's Missing Link," *Journal of Marketing*, vol. 46, no. 4, pp. 56–64, 1982, doi: 10.2307/1251362.

[13] A. F. Firat and C. J. Shultz, "From segmentation to fragmentation: Markets and marketing strategy in the postmodern era," *European Journal of Marketing*, vol. 31, no. 3/4, pp. 183–207, Apr. 1997, doi: 10.1108/EUM0000000004321.

[14] A. Jenkinson, "Beyond segmentation," *Journal of targeting, measurement and analysis for marketing*, vol. 3, no. 1, pp. 60–72, 1994.

[15] B. J. Jansen, K. Moore, and S. Carman, "Evaluating the performance of demographic targeting using gender in sponsored search," *Information Processing & Management*, vol. 49, no. 1, pp. 286–302, 2013.

[16] B. J. Jansen, J. O. Salminen, and S.-G. Jung, "Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics," *Data and Information Management*, vol. 4, no. 1, pp. 1–17, Mar. 2020, doi: 10.2478/dim-2020-0005.

[17] M. K. Saggi and S. Jain, "A survey towards an integration of big data analytics to big insights for value-creation," *Information Processing & Management*, vol. 54, no. 5, pp. 758–790, 2018.

[18] J. An, H. Kwak, J. Salminen, S. Jung, and B. J. Jansen, "Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data," *ACM Transactions on the Web (TWEB)*, vol. 12, no. 3, 2018.

[19] B. Jansen, J. Salminen, S. Jung, and K. Guan, *Data-Driven Personas*, 1st ed., vol. 14. Morgan & Claypool Publishers, 2021. Accessed: Feb. 10, 2021. [Online]. Available: https://www.morganclaypool.com/doi/abs/10.2200/S01072ED1V01Y202101HCI048

[20] R. Agarwal and V. Dhar, "Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research," *Information Systems Research*, vol. 25, no. 3, pp. 443–448, Sep. 2014, doi: 10.1287/isre.2014.0546.

[21] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from Social Big Data: Implications for Smart Tourism Destinations," *Information Processing & Management*, 2017.

[22] J. Salminen, M. Milenković, and B. J. Jansen, "Problems of Data Science in Organizations: An Explorative Qualitative Analysis of Business Professionals' Concerns," Dubai, Dec. 2017.

[23] H. Assael and A. M. Roscoe Jr, "Approaches to market segmentation analysis," *The Journal of Marketing*, pp. 67–76, 1976.

[24] C. Marcus, "A practical yet meaningful approach to customer segmentation," *Journal of Consumer Marketing*, vol. 15, no. 5, pp. 494–504, Oct. 1998, doi: 10.1108/07363769810235974.

[25] eMarketer, "Google and Facebook Tighten Grip on US Digital Ad Market - eMarketer," Sep. 21, 2017. https://www.emarketer.com/Article/Google-Facebook-Tighten-Grip-on-US-Digital-Ad-Market/1016494 (accessed Oct. 04, 2018).

[26] C. R. Goeldner, "Automation in marketing," *Journal of Marketing*, vol. 26, no. 1, pp. 53–56, 1962.

[27] G. W. Head, "What does automation mean to the marketing man?," *Journal of Marketing*, vol. 24, no. 4, pp. 35–37, 1960.

[28] J. Salminen, V. Yoganathan, J. Corporan, B. J. Jansen, and S.-G. Jung, "Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type," *Journal of Business Research*, vol. 101, pp. 203–217, Aug. 2019, doi: 10.1016/j.jbusres.2019.04.018.

[29] B. J. Jansen and S. Schuster, "Bidding on the buying funnel for sponsored search and keyword advertising," *Journal of Electronic Commerce Research*, vol. 12, no. 1, pp. 1–18, 2011.

[30] S. Khenissi and O. Nasraoui, "Modeling and counteracting exposure bias in recommender systems," *arXiv preprint arXiv:2001.04832*, 2020.

[31] B. J. Jansen, S. Jung, and J. Salminen, "Creating Manageable Persona Sets from Large User Populations," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2019, p. LBW2713:1-LBW2713:6. doi: 10.1145/3290607.3313006.

[32] J. Kotrlik and C. Higgins, "Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research," *Information technology, learning, and performance journal*, vol. 19, no. 1, p. 43, 2001.

[33] S. Thirumuruganathan, S. Jung, D. R. Robillos, J. Salminen, and B. J. Jansen, "Forecasting the nearly unforecastable: why aren't airline bookings adhering to the prediction algorithm?," *Electronic Commerce Research*, pp. 1–28, 2021, doi: https://doi.org/10.1007/s10660-021-09457-0.