International Journal of Innovative Research in Technology and Science

ISSN: 2321-1156 Volume XII, Issue I, Jan 2024

Survey on Social Media Analysis for Advancement of Digital Marketing

Ravi kumar [1], Sarwesh Site [2],

[1,2] Department of Computer Science & Engineering,

All Saints' College of Technology, Bhopal, India

Abstract

The research presents a survey conducting an extensive examination of the role and impact of social media analysis in the progression of digital marketing strategies. The survey explores various aspects of social media analytics, including data mining techniques, sentiment analysis, and trend identification, within the context of optimizing digital marketing efforts. The survey encompasses a comprehensive overview of current methodologies, tools, and technologies employed in social media analysis for marketing purposes. By critically evaluating existing practices and highlighting emerging trends, the paper provides valuable insights for marketers and researchers alike, aiming to contribute to the ongoing advancement of digital marketing strategies through informed and data-driven decision-making.
Keywords: Signal Analysis, Electrocardiogram, Machine Learning, Deep Learning

1 Introduction

In the dynamic landscape of contemporary business, the fusion of marketing strategies with social media data analysis has emerged as a powerful catalyst for success. As the digital sphere continues to evolve, organizations are recognizing the unparalleled potential of leveraging social media data to refine and optimize their marketing approaches [1]. This convergence not only marks a paradigm shift in the way brands interact with their audience but also signifies a strategic imperative for staying competitive in an increasingly interconnected world [2].

Social media platforms have evolved beyond mere communication channels; they now serve as vast repositories of valuable data reflecting user behaviors, preferences, and sentiments. Harnessing this wealth of information through advanced data analysis techniques opens up new avenues for marketers to tailor their campaigns with precision, resonate with target audiences, and drive impactful engagement [3]. The symbiotic relationship between marketing and social media data analysis stands poised to redefine how businesses understand, connect with, and influence their consumer base [4]. Deceptive reviews, commonly known as fake reviews, pose a significant challenge for online consumers as the prevalence of online marketplaces has given rise to a surge in fraudulent reviews. These misleading evaluations are frequently employed to attract or dissuade potential customers [5].

This introduction sets the stage for a deeper exploration into the multifaceted realm of marketing using social media data analysis. We delve into the transformative potential of data-driven insights, examining how organizations can navigate this intricate landscape to not only decode the intricacies of consumer behavior but also to craft marketing strategies that resonate authentically in the digital age. As we navigate this evolving landscape, the integration of social media data analysis becomes more than a trend; it becomes an integral component of a forward-thinking marketing arsenal, driving innovation, and ensuring relevance in an era defined by connectivity and digital influence [6].

2 Background and Related Work

The analytical data analysis and survey is divided into four sections, each corresponding to the primary steps of the process:

1. **data collection**, involves obtaining information about the users upon which personas are constructed.
2. **data enrichment**, involves examining the data collected earlier to extract additional insights.
3. **clustering**, involves identifying groups of users with similar characteristics based on the information gathered in the preceding two steps.
4. **persona generation**, involves crafting a persona for each recognized group.

Each section will delve into the progress made thus far, with a particular focus on recent advancements in utilizing enriched social media data. Additionally, the ongoing research challenge of fully automating the entire process is addressed.

2.1 Data collection

In this section, the principal data sources and methodologies employed for persona creation are presented in a chronological sequence. The discussion begins with hands-on qualitative methods, progresses to surveys, and concludes with the most recent advancements involving web data.

2.1.1 Qualitative methods

Traditionally, data collection relied on purely qualitative methods, which can be categorized into two types: implicit and explicit. Implicit methods, characterized by the principle of "don't ask, observe," encompass observations, field studies, and, to some extent, usability tests [7–9]. On the other hand, explicit methods involve directly posing questions to users, with interviews and focus groups being the most commonly used ones [7–9]. Whether implicit or explicit, these interactions need to be carefully designed, typically by experts in the relevant research field, to yield valuable insights for subsequent analysis.

While qualitative methods with their open-ended nature enable in-depth exploration of user behavior, they also come with certain limitations:

- **small coverage**: Typically, research is conducted with a limited number of users, often ranging from 10 to 20 individuals [7]. As a result, there is no assurance that personas accurately reflect the entire user base [10].
- **subjectivity**: The outcomes of interviews and other qualitative data are inherently subjective and may be influenced by biases from both users and researchers [7].

2.1.2 Surveys

Considering the aforementioned limitations, the concept of data-driven personas was introduced as a means to ground personas in larger-scale, real-world data by incorporating quantitative research methods alongside or in lieu of qualitative ones [7]. Surveys have emerged as a popular choice for implementing this approach, with 47% of papers on data-driven persona development reporting their utilization [11]. Surveys offer ease of sharing, facilitate a comprehensive exploration of the user base, and are simpler to cluster compared to the results of open-ended interviews.

One of the pioneers in employing surveys for persona development was R. Sinha in 2003, who utilized them when creating personas for a restaurant recommendation system [12]. Sinha and her team identified 32 relevant dimensions of the restaurant experience and crafted a survey, asking users to rate these dimensions on a five-point Likert scale. A similar approach was adopted by McGinn and Kotamraju in 2008 [13], where they collaborated with stakeholders to design a survey and obtained 1300 responses, a significantly larger dataset than what qualitative methods could achieve.

The use of surveys was further explored by Tu et al. in 2010 [14]. They proposed four steps for the persona creation process, with the initial step emphasizing, "*gather user data (personal data, users' relationships with the product and users' goals and motivations) through large scale questionnaires and user surveys.*" They underscored the critical role of questionnaire design in obtaining relevant dimensions for clustering and suggested a focus on capturing user behaviors, goals, and motivations. While they used a template designed by George Olsen in his article "*Persona creation and usage toolkit*" [9] to build their questionnaire, they noted the difficulty of fully adapting it to their specific situation, implying the absence of a universally applicable template.

2.1.3 Web data

The subsequent evolution in data collection emerged with the widespread adoption of web data. Approximately 30% of papers on data-driven personas report its utilization, and this percentage is on the rise, surpassing the use of surveys [11]. Web data encompasses sources such as social media platforms (e.g., YouTube [15]), online discussion forums [16], and online analytics (e.g., clickstream data [17]). The growing prevalence of web data has resulted in more extensive datasets and survey samples, with an average size approximately five times that of the initial surveys [11].

An illustrative example of leveraging online analytics was presented by Zhang et al. in 2016 [17]. Clickstream data, defined as sequences of users' clicks (or taps) while using a product, offers several advantages as a data source. Clickstreams directly depict user behavior and workflow within a product, are collected automatically without requiring human intervention, and their continual accumulation allows for easy updates to personas in case user workflows evolve over time. This approach proves particularly beneficial when personas need to encapsulate the user experience, such as in e-commerce websites or smartphone applications.

Social media is currently being explored as a source of both demographic and behavioral data, presenting a viable option due to the growing number of individuals with social media accounts who actively share content online. This content often provides insights into users' interests, opinions, and demographics, exhibiting notable similarities to information found in personas [18]. Additionally, most social media platforms offer APIs that enable the programmable collection of data on a large scale. However, researchers utilizing social media for persona creation must be mindful that the data within user profiles may not always reflect the truth, and a substantial number of accounts may be fake or bots (Twitter, for instance, deactivated up to 70 million suspicious accounts in 2018 [19]). Furthermore, privacy concerns are more pertinent than ever, necessitating full compliance with GDPR when

ISSN: 2321-1156 Volume XII, Issue I, Jan 2024

handling user data.

Predominantly led by Dr. Jim Jansen's team at the Qatar Computing Research Institute, research in this field focuses on the development of *APG (Automatic Persona Generation)*[1], a service that concentrates on automatically generating personas for a YouTube channel using YouTube analytics as the primary source, including data from comments, likes, and view counts categorized by content type and demographics [15]. Other online tools offering similar services emphasize online and social media analytics[2]. Consequently, users of such services are required to have an established online presence (e.g., website or social media accounts) and an analytics tool for monitoring related data (e.g., Twitter Analytics, Facebook Analytics, Google Analytics).

The use of more conventional social networks (beyond YouTube, which specializes predominantly in video content) has been explored to a lesser extent, with limited documentation available. An et al., during the development of personas for Al Jazeera[3], reported the use of Facebook and Twitter data but with certain limitations [20]. Facebook data was restricted to URLs shared by users who followed or discussed Al Jazeera's Facebook page, as the Facebook API requires explicit user consent to access any personal data. The use of Twitter data was also limited to users' biographies, aiming to extract non-behavioral aspects such as occupation and hobbies.

2.2 Data enrichment

Documentation on data enrichment is limited in persona literature, and it became a necessary step primarily with the introduction of social media as a data source. This is because information such as demographics and behavior was not directly accessible but had to be extracted from user profiles and posts [11].

Numerous studies aim to infer user demographics from social media, especially in academic research for examining gender disparities in publications. The most successful results are often achieved in predicting gender and age [21]. Various tools and APIs are available for such purposes, primarily utilizing names or images to predict a user's probable demographics. However, it's crucial to recognize that these are predictions and should be used cautiously to avoid introducing inaccurate data into the final personas. Whenever feasible, such predictions should undergo cross-validation using multiple data sources.

In determining user behavior, researchers commonly focus on inferring personality traits. Dos Santos et al. pioneered this approach by incorporating the Big Five personality framework in their surveys to create a behavioral persona for a pet robot [22]. Since then, significant research has been conducted to extract personality from social media user profiles. Initial approaches solely relied on public metrics, such as the number of friends, but proved inaccurate since users tend to have a large number of friends and followers regardless of their personality. Subsequent studies focused on examining the way users interact with their friends, incorporating factors such as the frequency and intensity of these interactions [23, 24].

Another valuable source of information is the content within social media posts. Analyzing the text or multimedia content of a user's posts enables one not only to comprehend the author's interests, values, and opinions on a specific topic or product but also to discern their preferred language, the times they are most active, and their tone of voice—factors that are all significant for defining a persona. Various Natural Language Processing (NLP) techniques are available to extract such insights from text. Semantic analysis, for instance, can identify entities such as nouns and corresponding adjectives [25]. Topic analysis can either extract or assign topics, revealing users' topical interests [26]. Sentiment scoring aids in understanding a user's attitude towards a specific topic [26, 27]. Deep learning solutions can also be employed to achieve similar results for images and videos [28]. It is crucial to recognize the significance of this, especially as multimedia content is rapidly gaining popularity over text. According to Mark Zuckerberg, CEO and founder of Facebook, "*Most of the content 10 years ago was text, and then photos, and now it's quickly becoming videos*" [4].

2.3 Clustering

Clustering, also referred to as segmentation, holds significant importance in the persona creation process as its output determines the number of personas and the characteristics that distinguish them.

Traditional segmentation involved manual analysis, where researchers examined the results of data collection to identify patterns and common themes. This method is highly subjective and often relies on researchers relying on their instincts, particularly if they lack expertise in the studied field [7].

The introduction of quantitative research in persona creation brought about more objective and algorithmic methods. Initially, dimensionality reduction techniques were utilized to unveil latent patterns in survey responses. Subsequent works adopted various clustering algorithms, selecting them based on the specific scenario for which personas were being created. Both of these approaches are detailed in the following subsections.

[1] persona.qcri.org

[2] www.delve.ai, www.mnemonic.ai

[3] www.aljazeera.com

[4] www.fastcompany.com/3057024/mark-zuckerberg-soon-the-majority-of-co

2.3.1 Dimensionality reduction

Surveys often contain a multitude of questions, leading to a large number of answers, and there may be latent correlations between them. In such cases, it can be beneficial to reduce the number of questions while retaining the majority of insights they provide. One method to achieve this is exploratory factor analysis, a statistical technique aimed at identifying underlying relationships between measured variables. However, this method has been documented only once in persona literature by McGinn and Kotamraju in 2008 [13].

A more popular and well-documented approach is the use of PCA (Principal Component Analysis) [11, 12, 14]. PCA is a dimension-reducing algorithm employed to extract information by eliminating non-essential elements with relatively fewer variations. While it was initially used independently [12, 14], recent research leveraging PCA often combines it with clustering algorithms, as presented in the following section [11], to mitigate the so-called curse of dimensionality [5].

2.3.2 Clustering algorithms

Clustering algorithms receive a set of data points as input and output a cluster index assigned to each of them. Below are some of the most commonly used clustering algorithms in persona creation.

K-Means An unsupervised machine learning algorithm that divides data into k clusters, where k is a user-chosen parameter. Its objective is to identify clusters so that data points within the same cluster are similar, while those in different clusters are more distant. It's important to note that similarity is determined by a distance metric, often Euclidean distance, though custom metrics may be necessary for complex problems or when dealing with a mix of categorical and numerical data. Additionally, as the initialization is random, results are not reproducible, although this can be addressed with an initialization setting. In persona creation, a drawback is the uncertainty regarding the optimal number of personas k, necessitating the use of other methods to determine it. Nevertheless, it remains the most widely used clustering algorithm in persona literature [11].

Non-negative matrix factorization (NMF) A matrix factorization method that approximately factors a matrix V (n x m) into two matrices W (n x r) and H (r x m), with the property that all three matrices have non-negative elements. Typically, r is chosen to be smaller than m to achieve dimensionality reduction. NMF is commonly used when dealing with user-content interactions. An et al. applied NMF to extract behavioral patterns from interactions between customer segments and YouTube videos [29]. In their case, matrix V had customer segments as rows and videos as columns, with each cell V_{ij} containing the view count of customer group i on video j. After factorization, they extracted video consumption patterns from matrix H and associated customer segments with each pattern through matrix W.

Hierarchical clustering (HC) An unsupervised machine learning algorithm that produces a hierarchical order of clusters arranged as a tree. It is an alternative to K-Means, providing reproducible results without requiring prior knowledge of the k parameter. However, it tends to be inefficient on large datasets in terms of speed and memory usage and is effective only when there is an underlying hierarchical structure in the data.

Other clustering algorithms Q-SIM and LDA. Q-SIM (Quality Similarity Clustering) was proposed as an alternative to K-Means clustering, replacing the hyperparameter k with a number representing the desired intra-cluster similarity degree [30]. Although it outperformed K-Means in a case study, its adoption has been limited in subsequent years [22], and further implementation support is lacking. LDA (Latent Dirichlet Allocation) is an NLP technique employed to uncover latent topics in text [31], making it suitable for clustering textual documents.

2.4 Persona generation

In this final step, the task involves constructing the ultimate persona profiles, one for each cluster. Regardless of the methods employed in the preceding steps (qualitative, quantitative, or mixed), the approach has remained consistent over time. Since each cluster possesses distinctive characteristics that set it apart from others, this step essentially entails "bringing clusters to life" by giving them an identity [7]. Typically, this is achieved by providing a name, photo, demographics (if not already present in the clusters), biography, quotes, and stories [7]. It's crucial to note that since personas are fictional, such characteristics should not be directly derived from the attributes of specific users. Instead, they should be abstracted from the data itself or generated from scratch. When doing so, care must be taken to ensure coherent attributes for believable personas, such as avoiding assigning a male name to a female persona [32].

This step posed challenges when attempting to fully automate the persona creation process, a problem addressed by the APG research team discussed in Section 2.1.3. They identified some challenges: firstly, when populating their personas with user quotes, they observed instances of inappropriate comments being selected, which negatively impacted the overall persona quality as assessed by marketing experts [33]. Secondly, to make personas believable, appropriate names corresponding to demographics need to be assigned. In response, they proposed a tool that takes age, country of origin, and gender as input and provides a suitable name [32]. The same consideration applies to

[5] `en.wikipedia.org/wiki/Curse_of_dimensionality`

ISSN: 2321-1156 Volume XII, Issue I, Jan 2024

photos: their initial approach involved purchasing a set of stock photos covering various demographics, but their recent works explore the possibility of using AI-generated images [34]. Finally, if personas are given a textual biography, it should align coherently with the rest of the data. As automatic persona creation research is relatively recent, the challenge of fully automating this last step is acknowledged as an open research problem [11, 33].

3 Proposed Approach

In response to the burgeoning significance of social media analysis in shaping digital marketing strategies, our proposed approach focuses on a comprehensive investigation into the evolving landscape of social media analytics. Emphasizing key elements such as data mining techniques, sentiment analysis, and trend identification, our research seeks to delve deeper into the practical applications of these methodologies within the realm of optimizing digital marketing efforts. Our approach involves a meticulous examination of current methodologies, tools, and technologies utilized in social media analysis for marketing purposes. By integrating a critical evaluation of existing practices and shedding light on emerging trends, our research aims to provide marketers and researchers with a robust framework for leveraging social media analytics in a manner that contributes to the continual evolution and enhancement of digital marketing strategies.

4 Conclusion and Future Work

This survey encapsulates a thorough exploration of the multifaceted role and impact of social media analysis on the progression of digital marketing strategies. By navigating through various facets of social media analytics, from data mining to sentiment analysis and trend identification, the research offers a holistic understanding of the tools and technologies employed in current practices. The critical evaluation of these methodologies not only provides insights into their efficacy but also sets the stage for the identification of emerging trends. Through this comprehensive overview, our paper aims to empower marketers and researchers, fostering an environment of informed decision-making based on data-driven approaches. Ultimately, the research contributes to the ongoing advancement of digital marketing strategies by providing valuable insights that align with the dynamic landscape of social media analytics.

References

[1] H. N. Bhor, T. Koul, R. Malviya, and K. Mundra, "Digital media marketing using trend analysis on social media," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, 2018, pp. 1398–1400. [doi: https://doi.org/10.1109/ICISC.2018.8399038

[2] Z. Aihua and C. Xi, "A review of social media and social business," in *2012 Fourth International Conference on Multimedia Information Networking and Security*, 2012, pp. 353–357. [doi: https://doi.org/10.1109/MINES.2012.44

[3] A. Singh, J. Singh, and A. Ghosal, "Impact of social media on stock market- a case of sentiment analysis using vader," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, vol. 1, 2022, pp. 300–304. [doi: https://doi.org/10.1109/COM-IT-CON54601.2022.9850668

[4] M. Jha, S. Khatiwada, and L. D. Li, "A conceptual framework to enhance business performance using social media: An australian context," in *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2021, pp. 1–6. [doi: https://doi.org/10.1109/CSDE53843.2021.9718492

[5] A. Mewada, R. K. Dewang, P. Goldar, and S. K. Maurya, "Sentibert: A novel approach for fake review detection incorporating sentiment features with contextual features," in *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*, ser. IC3-2023. Association for Computing Machinery, 2023, pp. 230–235. [doi: https://doi.org/10.1145/3607947.3607991

[6] M. Gupta, R. Kumar, A. Sharma, and A. S. Pai, "Impact of ai on social marketing and its usage in social media: A review analysis," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–4. [doi: https://doi.org/10.1109/ICCCNT56998.2023.10308092

[7] S. Mulder and Z. Yaar, "Approaches to creating personas," *The user is always right: A practical guide to creating and using personas for the web. Berkeley, CA: New Riders*, pp. 33–54, 2007.

[8] J. Pruitt and J. Grudin, "Personas: practice and theory," in *Proceedings of the 2003 conference on Designing for user experiences*, 2003, pp. 1–15.

[9] G. Olsen, "Persona creation and usage toolkit," *Retrieved March*, vol. 25, p. 2014, 2004.

[10] C. N. Chapman and R. P. Milham, "The personas' new clothes: methodological and practical arguments against a popular method," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 5. SAGE Publications Sage CA: Los Angeles, CA, 2006, pp. 634–636.

[11] J. Salminen, K. Guan, S.-G. Jung, and B. J. Jansen, "A survey of 15 years of data-driven persona development," *International Journal of Human–Computer Interaction*, pp. 1–24, 2021.

ISSN: 2321-1156 Volume XII, Issue I, Jan 2024

[12] R. Sinha, “Persona development for information-rich domains,” in *CHI’03 extended abstracts on Human factors in computing systems*, 2003, pp. 830–831.

[13] J. McGinn and N. Kotamraju, “Data-driven persona development,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2008, pp. 1521–1524.

[14] N. Tu, Q. He, T. Zhang, H. Zhang, Y. Li, H. Xu, and Y. Xiang, “Combine qualitative and quantitative methods to create persona,” in *2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering*, vol. 3. IEEE, 2010, pp. 597–603.

[15] J. An, H. Kwak, S. Jung, J. Salminen, M. Admad, and B. Jansen, “Imaginary people representing real numbers: Generating personas from online social media data,” *ACM Transactions on the Web (TWEB)*, vol. 12, no. 4, pp. 1–26, 2018.

[16] J. Huh, B. C. Kwon, S.-H. Kim, S. Lee, J. Choo, J. Kim, M.-J. Choi, and J. S. Yi, “Personas in online health communities,” *Journal of biomedical informatics*, vol. 63, pp. 212–225, 2016.

[17] X. Zhang, H.-F. Brown, and A. Shankar, “Data-driven personas: Constructing archetypal users with clickstreams and user telemetry,” in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 5350–5359.

[18] A. Humphrey, “User personas and social media profile,” *Persona Studies*, vol. 3, no. 2, pp. 13–20, 2017.

[19] B. News, “Twitter ’shuts down millions of fake accounts’,” https://www.bbc.com/news/technology-44682354, July 2018, accessed: 14 May 2021.

[20] J. An, H. Cho, H. Kwak, M. Z. Hassen, and B. J. Jansen, “Towards automatic persona generation using social media,” in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*. IEEE, 2016, pp. 206–211.

[21] N. Cesare, C. Grant, Q. Nguyen, H. Lee, and E. O. Nsoesie, “How well can machine learning predict demographics of social media users?” *arXiv preprint arXiv:1702.01807*, 2017.

[22] T. F. dos Santos, D. G. de Castro, A. A. Masiero, and P. T. A. Junior, “Behavioral persona for human-robot interaction: a study based on pet robot,” in *International Conference on Human-Computer Interaction*. Springer, 2014, pp. 687–696.

[23] S. Adali and J. Golbeck, “Predicting personality with social behavior,” in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2012, pp. 302–309.

[24] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, “Predicting personality from twitter,” in *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 2011, pp. 149–156.

[25] D. H. Maulud, S. R. Zeebaree, K. Jacksi, M. A. M. Sadeeq, and K. H. Sharif, “State of art for semantic analysis of natural language processing,” *Qubahan Academic Journal*, vol. 1, no. 2, pp. 21–28, 2021.

[26] S. Huang, W. Peng, J. Li, and D. Lee, “Sentiment and topic analysis on social media: a multi-task multi-label classification approach,” in *Proceedings of the 5th annual ACM web science conference*, 2013, pp. 172–181.

[27] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, “Sentiment analysis on social media,” in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2012, pp. 919–926.

[28] P. Rodríguez, D. Velazquez, G. Cucurull, J. M. Gonfaus, F. X. Roca, S. Ozawa, and J. Gonzàlez, “Personality trait analysis in social networks based on weakly supervised learning of shared images,” *Applied Sciences*, vol. 10, no. 22, p. 8170, 2020.

[29] J. An, H. Kwak, S.-g. 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, pp. 1–19, 2018.

[30] A. A. Masiero, R. de Carvalho Destro, O. A. Curioni, and P. T. A. Junior, “Automa-persona: A process to extract knowledge automatic for improving personas,” in *International Conference on Human-Computer Interaction*. Springer, 2013, pp. 61–64.

[31] D. Bamman, B. O’Connor, and N. A. Smith, “Learning latent personas of film characters,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2013, pp. 352–361.

[32] S.-G. Jung, J. Salminen, and B. J. Jansen, “All about the name: Assigning demographically appropriate names to data-driven entities,” in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021, p. 4034.

[33] J. Salminen, S.-g. Jung, and B. J. Jansen, “The future of data-driven personas: A marriage of online analytics numbers and human attributes.” in *ICEIS (1)*, 2019, pp. 608–615.

[34] J. Salminen, S.-g. Jung, A. M. S. Kamel, J. M. Santos, and B. J. Jansen, “Using artificially generated pictures in customer-facing systems: an evaluation study with data-driven personas,” *Behaviour & Information Technology*, pp. 1–17, 2020.