Dec 11th, 12:00 AM

MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data

Daniel Konadl
University of Regensburg, daniel.konadl@wiwi.uni-regensburg.de

Janik Wörner
University of Regensburg, janik.woerner@ur.de

Isabel Maria Schmid
University of Regensburg, isabel.schmid@wiwi.uni-regensburg.de

Susanne Leist
University of Regensburg, susanne.leist@ur.de

MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data

Completed Research Paper

Janik Wörner
University of Regensburg
Universitätsstraße 31,
93053 Regensburg, Germany
Janik.Woerner@ur.de

Daniel Konadl
University of Regensburg
Universitätsstraße 31,
93053 Regensburg, Germany
Daniel.Konadl@ur.de

Isabel Schmid
Regensburg, Germany
Isabel.m.schmid@gmail.com

Susanne Leist
University of Regensburg
Universitätsstraße 31,
93053 Regensburg, Germany
Susanne.Leist@ur.de

Abstract

The early identification of new and auspicious ideas leads to competitive advantages for companies. Thereby, topic modeling can serve as an effective analytical approach for the automated investigation of trends from unstructured social media data. However, existing trend analysis tools do not meet the requirements regarding (a) Product Development, (b) Customer Behavior Analysis, and (c) Market-/Brand-Monitoring as reflected within extant literature. Thus, based on the requirements for each of these common marketing-related use cases, we derived design principles following design science research and instantiated the artifact "MANTRA" (MArketiNg TRend Analysis). We demonstrated MANTRA on a real-world data set (~1.03 million Yelp reviews) and hereby could confirm remarkable trends of vegan and global cuisine. In particular, the importance of meeting all specific requirements of the respective use cases and especially flexibly incorporating several external parameters into the trend analysis is exemplified.

Keywords: Social Media Analytics, Trend Analysis, Topic Modeling-Based Tool, Design Science, Marketing

Introduction

In the fast-paced and competitive world of business today, it is crucial for companies to keep track of trends (e.g., Endo et al. 2015; Lozano et al. 2017). Trends are complex, multidimensional phenomena that evolve over time and that can exert a significant impact on society in case they prevail (cf. Burmeister et al. 2002). In this context, social media plays a pivotal role as it allows trendsetters to openly discuss emerging trends across various industries (e.g., technology, fashion, or food trends) (e.g., Jeong et al. 2019; Park et al. 2021) and enables customers to express their unfiltered opinions (Yan et al. 2014). Thus, based thereon, companies can get an impression of what customers discuss as trends and what they think that future products and services should look like. In this respect, there are three focal points in the extant marketing research literature that deal with identifying and implementing trends: Developing a sound comprehension of feasible development paths of products and services (Puranam et al. 2017), customers' cultural norms and behaviors (Humphreys and Wang 2018), and market structures (Netzer et al. 2012). Our research considers these focal points by three different use cases and aims to support the early exposure of trends.

To identify trending topics from social media, marketing departments can apply different automated analysis techniques (cf. Batrinca and Treleaven 2015; Fan and Gordon 2014; Stieglitz et al. 2018). There are quantitative approaches that identify trends based on the frequency by which certain keywords appear within social media posts over time. However, this is insufficient, as trends, whose constituent elements are not necessarily known in advance, cannot be found in this way. Supervised learning could also be applied to trend analysis, but this requires ex-ante knowledge about trend characteristics and trend impulses. Unsupervised learning techniques can autonomously extract disruptive and previously unknown trends from large text volumes. Trending topics can thereby include combinations of two or more words that are not unusually topical or frequently mentioned together (Khan et al. 2021; Reisenbichler and Reutterer 2019). In addition, identifying meaningful trends can benefit from previously gained domain knowledge (cf. Burmeister et al. 2002). Therefore, semi-supervised topic modeling turns out to be a suitable automated trend analysis technique as it determines latent relationships between the individual components of a trend (Crain et al. 2012; Hong and Davison 2010) and allows its users to provide seed information as semantic anchors to guide the topics to be discovered (Jagarlamudi et al. 2012).

The potential that automated trend analysis techniques propose has already been recognized in practice and science. Nevertheless, available solutions show remarkable drawbacks, as they do not cover important Design Requirements (DRs) that are deemed crucial for a trend analysis concerning the common marketing-related use cases (e.g., incorporation of different geolocation data and customer characteristics as external parameters) that we could derive. Hence, the capabilities for comprehensively assessing trends from different perspectives (i.a., products, customers, markets) are restricted. Companies need to operate different trend analysis approaches to adequately support these common marketing-related use cases, leading, i.a., to the burden of tediously compiling individual results to gain a holistic view of trends. Moreover, the available tools proposed for automated trend analysis of marketing-related use cases do not offer marketing representatives to incorporate prior domain knowledge. As a result, the extracted trends are not necessarily semantically tailored to the company and its fields of activity, which severely impedes the integration of the gained insights into decision-making processes within marketing departments.

We address these drawbacks of existing solutions and propose the design and a concrete instantiation of a software tool as an automated artifact for trend analysis. By including a topic modeling technique that particularly enables the incorporation of previously gained domain knowledge, we show the proficiency of this technique in automated trend analysis. Regarding the DRs that a corresponding automated trend analysis tool needs to meet for common marketing-related use cases, we focus especially on the combination of different machine learning techniques, which leads to the trend analysis being purposeful. Against this background, the following two Research Questions (RQs) are posed:

- ***RQ1:*** *Which DRs should a topic modeling-based trend analysis tool meet that supports the marketing-related use cases Product Development, Customer Behavior Analysis, and Market-/Brand-Monitoring?*
- ***RQ2:*** *How can such a topic modeling-based trend analysis tool be designed and implemented, and which contributions can be derived?*

The remainder of this paper is structured as follows: the next section provides the conceptual foundations and the related work. Subsequently, the research procedure, following the Design Science (DS) approach (Hevner et al. 2004; Peffers et al. 2007), is described. The next section deals with the compilation of the DPs that we have derived from the DRs. Then, we report the technical realization of the tool, demonstrate its application on about 1.03 million Online Customer Reviews (OCRs), and present the resulting outcomes. The paper closes with concluding remarks, potential limitations, and future research ideas.

Conceptual Foundations and Related Work

Conceptual Foundations

Social media entails *"a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation of User Generated Content"* (Kaplan and Haenlein 2010, p.61). Online social platforms (e.g., Yelp) as a form of social media thereby serve as an important medium for the exchange between companies and customers by unstructured and opinionated content (e.g., textual social media posts). Trendsetters (e.g., domain luminaries or pioneers in certain fields) can share their thoughts on trending topics and discuss what future products and services should look like (e.g., Jeong et al. 2019; Park et al. 2021; Tucker and Kim 2011). In this respect, OCRs constitute *"peer-generated product evaluations posted on company or third party websites"* (Mudambi and Schuff, 2010, p.186), enabling users to evaluate companies' current products and services by disseminating their experiences. Thus, OCRs are also beneficial for companies as they include customers' expectations and opinions concerning products, services, or the company in general (cf. Hicks et al. 2012; Sigala 2012). As OCRs need to be online, social media platforms have shown effective for their dissemination in practice. Therefore, companies can harness OCRs to gain insights about central marketing-related dimensions unfiltered and in real-time (Yan et al. 2014): i.a., opinions about products and services (Ko et al. 2018; Moro et al. 2020), customers' behaviors (Humphreys and Wang 2018), and market structure and competition (Netzer et al. 2012; Tirunillai and Tellis 2014). Keeping track of the evolution of topics and these marketing-related dimensions is central to trend analysis. In this way, marketing departments can enhance the effectiveness of brand message placement and the allocation of appropriate resources to marketing campaigns depending on geographical and temporal developments (Hong et al. 2012; Lozano et al. 2017).

To reveal trending topics, marketing representatives can apply topic modeling, which refers to the use of generative probability models for determining latent relationships within a corpus of text data. Hereby, semi-supervised topic modeling allows the user to guide the topic discovery process by providing seed information as semantic anchors (Jagarlamudi et al. 2012). In this way, different contextual information (e.g., geolocation, time, user characteristics, a company's field of activity) can be incorporated to tailor the trends to be derived. Despite these potentials, semi-supervised topic modeling has yet not comprehensively been harnessed for trend analysis within marketing use cases.

Distinctive Use Cases of Topic Modeling-Based Trend Analysis

To identify common and distinctive marketing-related use cases of topic modeling-based trend analysis, we have comprehensively searched and consolidated the extant research literature. Thereby, we have aligned with the procedure for conducting a systematic literature review as proposed by Vom Brocke et al. (2015). In the **first step**, we applied a keyword-based search on *ACM Digital Library, AIS Electronic Library, EBSCOhost, Emerald Insight, Google Scholar, IEEE Xplore, SAGE Publications, SpringerLink, and the Wiley Online Library (keywords: "topic model", "topic modeling", "trend analysis", "trend tracking", "marketing", "social media", and several combinations thereof)*. Within the keyword-based search, we conducted iterative keyword refinements and could thereby derive the following further keywords: *"online customer review", "product development", "customer behavior", "customer behavior change", "market surveillance", "market monitoring", and "brand monitoring"*. We only kept papers dealing with topic modeling within the domain of marketing for automated trend analysis on unstructured social media data. Based thereon, the three involved researchers independently analyzed the title, abstract, and thematic focus of the initially derived *682 papers*. It turned out that there are papers that mention topic modeling but do not apply it to trend analysis. Some papers used social media and marketing-related terms, but these concepts were not reflected in the application of topic modeling *(remaining: 52)*. In the cases where assessments of the researchers showed ambiguities, the relevance and fit of the respective paper to our investigation were determined by clearing up the ambiguities through discussions and justifications among the involved researchers *(remaining: 38)*. Duplicates were eliminated, only papers written in English and only those that have passed a peer-review assessment were kept *(remaining: 26)*. In the **second step**, we conducted a forward and a backward search, resulting in another *3 relevant papers* that meet the defined inclusion criteria. Therefore, in total *29 papers* could be derived.

Subsequently, we have investigated the contents of the derived papers. Thereby, the wording of the statements in the examined sources often differed considerably, and the logical consolidation of the individual statements offered room for interpretation. Using qualitative content analysis and inductive category development, we could carry out an abstraction from the basic data, while, at the same time, deriving a coherent image of these data (cf. Mayring 2000; Mayring 2002). In this way, in particular, three distinctive marketing-related use cases could be identified for automated topic modeling-based trend analysis: (a) **Product Development** (e.g., Bae et al. 2018; Irawan et al. 2020; Jeong et al. 2019), (b) **Customer Behavior Analysis** (e.g., Bhor et al. 2018; Chen et al. 2012; Zhang et al. 2016) and (c) **Market-/Brand-Monitoring** (e.g., Lozano et al. 2017; Qu et al. 2015; Zhao et al. 2019).

Design Requirements for a Topic Modeling-Based Trend Analysis Tool

After having derived the distinctive marketing-related use cases, we next aimed to derive DRs for the respective use cases. Because different papers pointed to the same requirements but in (slightly) different wordings, we applied the qualitative content analysis (cf. Mayring 2000; Mayring 2002) here as well. In this way, we could systematically condense the available data and consolidate the contents that relate to the respective DRs. In total, we could derive 14 DRs that we will describe in more detail in the following:

By applying topic modeling for (a) **Product Development**, marketing representatives can develop an understanding of how customers perceive their products, services, and corresponding features. Topics identified from OCRs may, e.g., indicate that customers perceive the product packaging as unecological. Harnessing OCRs may bring up the idea of replacing the existing plastic packaging with paper packaging so that customers' expectations are subsequently met (cf. Park et al. 2021). As the prevailing literature shows for the use case (a), customers' product and service expectations can differ across geographical markets. To be successful and competitive in a targeted market, marketing representatives need to know which features to design in which way to meet the global and local customers' expectations. Therefore, topic modeling needs to **(DR1)** incorporate different geolocation data as an external parameter (e.g., Bae et al. 2018; Ding et al. 2020; Ha et al. 2017) that match the generated topics with the corresponding geographical locations. Customers' self-reported opinions of products and services also play an important role. Marketing representatives aim at retaining features evoking positive customer perceptions, while features evoking negative sentiments need to be improved. Thus, a trend analysis tool needs to **(DR2)** incorporate the sentiment of social media posts (e.g., Irawan et al. 2020; Jeong et al. 2019). Thereby, numerous opportunities for improving products and services can be determined, and certain features are given greater importance than others. Yet, it is not feasible to address all identified issues because of limited resources in marketing departments. Thus, it is necessary to support marketing representatives **(DR3)** to assess the product and service feature favorability (e.g., Ding et al. 2020; Moro et al. 2020; Tucker and Kim 2011). Regarding use case (a), e.g., Production Theory and Innovation Theory can serve as theoretical foundations. Innovation Theory proposes harnessing external and internal sources for identifying new and auspicious ideas of what future products may look like. These ideas are assessed and refined so that the most promising ideas are integrated as input factors into the product refinement and development process (cf. Axtell et al. 2000; Ye and Kankanhalli 2018). So, we argue that it is imperative for a trend analysis tool to support the refinement of the identified trends and to enable marketing representatives to **(DR4)** link the anticipated customer preferences with new features (e.g., Jeong et al. 2019; Ko et al. 2018; Tucker and Kim 2011).

In contrast to (a), the focus in (b) **Customer Behavior Analysis** is on the authors of the social media posts. Marketing representatives responsible for customer targeting must know how customer behaviors and associated trends differ for certain customer characteristics like age or gender (e.g., Bhor et al. 2018; Chen et al. 2012). Customer Behavior Analysis can for example reveal that older people perceive comfort as an essential product property. To increase favorability among older customers, a company may communicate its superiority regarding comfort to this respective target group (cf. Duncan and Moriarty 1998). For certain customer characteristics, there are as well different customer behaviors. As such, the needs for online shopping do distinguish between males (e.g., the accuracy of product descriptions) and females (e.g., fair pricing) (cf. Ulbrich et al. 2011). In this vein, Customer Focus Theory can be harnessed as it proposes companies to align their actions to customers' characteristics and expectations (Gulati and Oldroyd 2005; Tseng and Piller 2003). To establish the topic-person connections, topic modeling techniques, therefore, need to **(DR5)** incorporate different user-related information and customer characteristics (e.g., age, gender, group-related attitudes, and values) as external parameters (e.g., Bhor et al. 2018; Shi et al. 2018; Zhang et al. 2016). Based thereon, marketing representatives can tailor marketing messages with proposals that the individual customer groups appreciate most. As (b) is about customer behavior, **(DR6)** a functionality to filter posts is essential to exclude company posts before generating customer-related topics (Iwata et al. 2009; Saha and Sindhwani 2012). Fast reactions **(DR7)** are important regarding topic modeling techniques for this use case because customers' perceptions can change rapidly (Bhor et al. 2018; Iwata et al. 2009; Sasaki et al. 2014) and, while, identified concerns about products or services require a certain time for refactoring, customers need to be rapidly contacted and appeased.

The third use case (c) **Market-/Brand-Monitoring** is about investigating how brands and companies act and which strategies they deploy. Therefore, (c) is concerned with the companies and brands instead of (a), their products and services. Applying trend analysis to the social media posts of competitors could reveal that competing brands include reporting about their sustainable acting within public communications. To support decisions on the market-/brand level, the topics must be generated for the respective companies or brands. Beyond generating topics according to geographical markets, the topic modeling technique **(DR8)** must also be capable of incorporating brands as external parameters (e.g., Endo et al. 2015; Lozano et al. 2017; Qu et al. 2015). Based thereon, a company may, e.g., adapt its market strategy to focus on niche markets that other brands have yet paid little attention to. Innovation Theory (cf. Sundbo 1995) thereby serves as a foundation as it recommends how to successfully place new innovative products and services within a market. Hereby, the Strategic Innovation Paradigm points out that companies need to assess how competitors shape their strategic communications concerning innovations. Thus, a trend analysis tool **(DR9)** must support the identification of competitors and their market orientations (e.g., Lozano et al. 2017; Park et al. 2021; Zhao et al. 2019). Moreover, marketers need not only support in identifying emerging competitors but also in understanding the behaviors of current ones (e.g., communication patterns and discount strategies). Therefore, **(DR10**) filter criteria like time periods, geolocations, and brands should be proposed to narrow down the dataset (e.g., Sohn et al. 2019; Valdez et al. 2018; Zhao et al. 2019).

Beyond use case-specific DRs, we could identify DRs that apply to all three use cases. One theory posing fundamental implications across all use cases is Social Media Analytics Theory (e.g., Fan and Gordon 2014; Stieglitz et al. 2018). When applying social media analytics for trend analysis on unstructured social media data, a central necessity is **(DR11)** dealing with a huge number of social media posts because of the ever-increasing speed of social media posts emergence (e.g., Jeong et al. 2019; Lozano et al. 2017; Stieglitz et al. 2018). Further, it is essential to provide means to include **(DR12)** historical data so that past developments of trends can be considered when assessing the current state of a trending topic (e.g., Bhor et al. 2018; Lozano et al. 2017). Because trends constitute dynamic developments and not solely static points in time, the topic modeling technique needs to integrate different temporal parameters into the topic generation. Marketers require **(DR13)** support in assessing the course of topics over time at different levels of granularity and in identifying emerging and declining trends (e.g., Ha et al. 2017; Zhong and Schweidel 2020). In general, applying the results of topic modeling requires advanced technique-related knowledge. Thus, **(DR14)** the results need to be visualized (by charts and plots) so that marketing representatives are supported in making sense of the derived topics (Bae et al. 2018; Ha et al. 2017; Yang et al. 2017).

Use Cases	Design Requirements	Sources
(a) Product Development	**DR1.** Incorporation of different geolocation data as external parameters	Bae et al. 2018; Ding et al. 2020; Endo et al. 2015; Ha et al. 2017; Moro et al. 2020; Park et al. 2021
	DR2. Consideration of sentiment	Irawan et al. 2020; Jeong et al. 2019; Ko et al. 2018; Moro et al. 2020
	DR3. Providing means to assess the importance of a product or service feature	Bai et al. 2023; Ding et al. 2020; Jeong et al. 2019; Mastrogiacomo et al. 2021; Moro et al. 2020; Tucker and Kim 2011
	DR4. Linking of newly discovered preferences with new features	Axtell et al. 2000; Bai et al. 2023; Ding et al. 2020; Jeong et al. 2019; Ko et al. 2018; Puranam et al. 2017; Tucker and Kim 2011; Ye and Kankanhalli 2018
(b) Customer Behavior Analysis	**DR5.** Incorporation of different user-related information and customer characteristics as external parameters	Bhor et al. 2018; Chen et al. 2012; Gulati and Oldroyd 2005; Puranam et al. 2017; Qu et al. 2015; Shi et al. 2018; Tseng and Piller 2003; Zhang et al. 2016
	DR6. Provide means to filter out social media posts being posted by the company itself	Gulati and Oldroyd 2005; Iwata et al. 2009; Saha and Sindhwani 2012; Tseng and Piller 2003
	DR7. Support fast reactions to changing customer perceptions	Bhor et al. 2018; Iwata et al. 2009; Sasaki et al. 2014
(c) Market-/ Brand-Monitoring	**DR8.** Incorporation of different brands and geolocation data as external parameters	Endo et al. 2015; Lozano et al. 2017; Qu et al. 2015; Sohn et al. 2019; Zhao et al. 2019
	DR9. Identify differences of how enterprises communicate and place themselves in markets	Park et al. 2021; Sundbo 1995; Tirunillai and Tellis 2014; Valdez et al. 2018; Zhao et al. 2019
	DR10. Apply time periods, geolocations, and brands as filter criteria	Sohn et al. 2019; Sundbo 1995; Valdez et al. 2018; Zhao et al. 2019
Use Case Independent Design Requirements	**DR11.** Deal with large numbers of social media posts	Fan and Gordon 2014; Jeong et al. 2019; Lozano et al. 2017; Puranam et al. 2017; Stieglitz et al. 2018; Yan et al. 2015
	DR12. Capability of analyzing historical data	Bai et al. 2023; Bhor et al. 2018; Ha et al. 2017; Lozano et al. 2017; Puranam et al. 2017; Sohn et al. 2019; Zhong and Schweidel 2020
	DR13. Identification of trends over time (topic evolution) at different levels of granularity (e.g., phases of a day, days of a week, seasons)	Ha et al. 2017; Lozano et al. 2017; Puranam et al. 2017; Tucker and Kim 2011; Zhang et al. 2015; Zhong and Schweidel 2020
	DR14. Visualization of the results	Bae et al. 2018; Ha et al. 2017; Yang et al. 2017

Table 1. Identified Use Cases and Corresponding Design Requirements

Assessment of Tools for Automated Trend Analysis on Social Media

To assess the capabilities of automated trend analysis tools for unstructured social media data, we first identified corresponding tool providers. The Google search engine and social media analytics groups within online social networks (i.a., LinkedIn) were drawn upon for this purpose. Tools identifying trends based on ex-ante knowledge (i.a., approaches that assess trends by the frequency by which certain keywords appear) were disregarded, as this tremendously hinders the extraction of disruptive and previously unknown trends.

By three researchers conducting the identification of tool providers independently, the possibility of overlooking established tool providers could be reduced. The identified tool providers were then consolidated and potential disagreements about the tools were resolved. In this way, we could cover a broad spectrum of providers and take an up-close look at the most established tools for automated trend analysis *(Brandwatch, Hootsuite, IBM, Keyhole, Meltwater, MineMyText, NetBase Quid, Pulsar, Q-Markets, Symanto, Synthesio, Trend-Sonar)*. Based on the DRs that we could derive *(see Table 1)*, a tool survey was designed to assess whether the available trend analysis tools cover the posed requirements. The tool survey was validated by the involved researchers and subsequently applied to the data collection procedure. For this reason, we interviewed sales representatives from these tool providers on the specific features and additionally installed demo versions of the tools. Based on the interview results and by independently testing the functionalities of the trend analysis tools against the posed DRs, we could gain detailed and comprehensive insights into the capabilities as well as into the drawbacks of those trend analysis tools.

Solution Prov.	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
# Employees	~ 10	~ 30	~ 60	~ 70	~ 70	~ 120	~ 300	~ 1.3k	~ 1.4k	~ 2.2k	~2.5k	~ 290k
General Approach	NLP	NLP	AI	AI	NLP	NLP	NLP	AI	AI	NLP	NLP	AI
DR1	-	-	○	-	-	-	-	-	-	-	-	○
DR2	-	●	-	●	-	-	●	●	●	-	●	●
DR3	●	●	○	●	●	●	●	●	●	●	●	●
DR4	-	○	-	-	●	○	○	-	○	-	-	-
DR5	-	-	-	-	-	-	-	-	-	-	-	-
DR6	●	○	●	●	-	●	-	●	●	●	●	●
DR7	*DR7 was **not** included in the comparison because a fast execution is highly dependent on the underlying operating hardware.*											
DR8	-	-	○	-	-	-	-	-	-	-	-	-
DR9	-	-	-	○	○	○	●	-	-	-	-	●
DR10	●	●	●	●	●	●	●	●	●	●	●	●
DR11	●	●	●	●	●	●	●	●	●	●	●	●
DR12	●	○	●	●	●	●	●	●	●	●	●	●
DR13	○	○	○	○	○	○	○	○	○	○	○	○
DR14	●	●	●	●	●	●	●	●	●	●	●	●

Table 2. Results of Assessing Solutions for Automated Trend Analysis
AI: Artificial Intelligence (i.a., neural networks, support vector machines), **NLP:** Natural Language Processing (i.a., unsuperv. topic modeling)

● = applicable
○ = partly applicable

As our assessment could reveal, there is no automated tool that offers marketing representatives to unveil the latent relationships between the individual components of a trend and that enables them at the same time to incorporate their previously gained domain knowledge. The tools lack to meet various specific DRs of the marketing-related use cases (a)-(c), none of the tools thereby manages to cover all of these DRs, and, in particular, they cannot flexibly integrate the various external parameters at different granularity. Nonetheless, the perceptions of products and services differ across continents, countries, and federal states because of globally and locally differing customer expectations. Customer behaviors do differ for certain customer characteristics (e.g., age, gender, group-related attitudes, and values). Without the ability to flexibly integrate different external parameters like time (e.g., days of a week, phases of a day, seasons), geolocations, and user-related information at different granularity, an automated trend analysis tool does not sufficiently support trend analysis and the marketing-related use cases (a)-(c). Beyond the revealed drawbacks of the available tools, extant literature clearly points out the need of such a comprehensive topic modeling-based tool that supports all three use cases. With this research, we aim to close this gap.

Research Procedure

To accomplish the development of a tool for automated trend analysis regarding the common marketing-related use cases that we could derive from the extant research literature, we applied Design Science (DS) research. A widely recognized suggestion on how to conduct DS projects was introduced by March and Smith (1995) and Peffers et al. (2007). In this respect, DSR represents a synthesis of the activities "build/development" and "justify/evaluate" with the main goal of designing an IT artifact to address an organizational problem *(see Figure 1)* (cf. Gregor and Hevner 2013; Hevner et al. 2004; Peffers et al. 2007).

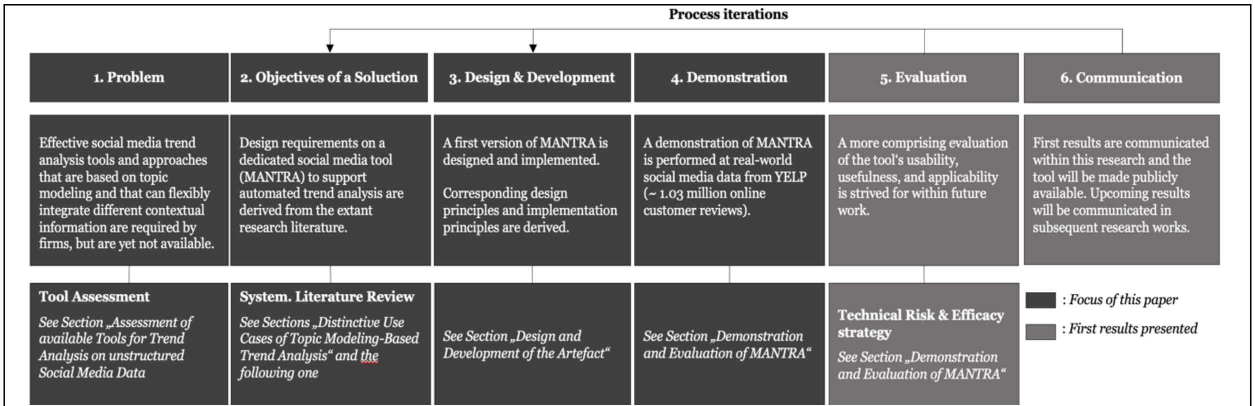

Figure 1. Procedure of the Research, own Representation Based on Peffers et al. (2007)

As a first step, **(1) corresponding problems** of existing software tools for trend analysis were identified *(see Section "Introduction" and Table 2)*. Automated tools for trend analysis neither do meet the requirements that are indispensable to the successful application of trend analysis nor do provide flexibly incorporating external parameters (e.g., by means of semi-supervised topic modeling). Thus, our **(2) objective** is to provide and combine a set of machine learning techniques, based on the DRs, to facilitate trend analysis regarding common marketing-related use cases based on semi-supervised topic modeling.

The third step contains the **(3) design and development** *(see Section "Design and Development of the Artifact")* of the solution. By conducting a systematic literature review (cf. Vom Brocke et al. 2015), we could identify DRs an automated trend analysis tool needs to meet to be comprehensive for common marketing-related use cases *(see Section "Distinctive Use Cases of Topic Modeling-Based Trend Analysis")*. To fill the gaps identified within step (1), we focus on the design of the technical realization of the tool by establishing DPs that are founded on the identified DRs. Thus, our approach was established to support a comprehensive trend analysis for marketing-related use cases and to eliminate the drawbacks of existing solutions.

Steps **(4)** and **(5)** deal with the **demonstration** and **evaluation** of the implemented artifact. To define a proper setting for rigorously demonstrating and evaluating the utility of the artifact, we related to Venable et al. (2016) who have defined four evaluation strategies for DSR projects. We aligned our DSR-project-specific evaluation strategy to the "Technical Risk & Efficacy strategy" because the focal aim of our investigation was to rigorously demonstrate that the evaluand "MANTRA" (MArketiNg TRend Analysis) achieves its purpose (i.e., identify meaningful and sound trends from unstructured social media data) and that the incorporation of semi-supervised topic modeling and meeting several DRs are thereby imperative (cf. March and Smith 1995; Peffers et al. 2012). Thus, we decided to apply one single summative evaluation episode (cf. Venable et al. 2016) in which trends that have already occurred are reproduced retrospectively. Thereby, the results produced by MANTRA are verified and confirmed with two trend reports. Beyond, we plan to investigate the tool within a controlled environment through a laboratory experiment where participants (i.a., experts from the food industry) identify the soundness of the trends and provide feedback on how to improve the tool before we will conduct a more extensive evaluation in a natural setting (e.g., within a field study). In step **(6)**, we aim to **document** and **communicate** the results of our research.

With this procedure, we align our research also with the guideline of Hevner et al. (2004). Regarding the design cycle, we present our artifact as the result that has gone through the process of demonstration (application to a Yelp dataset) and evaluation by a survey of an expert panel from the National Restaurant Association Board (NRAB) *(see Section "Demonstration and Evaluation of MANTRA")*. Regarding the relevance cycle, we identified several DRs from research literature and theory *(see Table 1)* that guided the design of the artifact, and so the practical application of our artifact brought up several contributions for practice *(see Section "Discussion and Contributions")*. In view of the rigor cycle, we used several techniques (i.a., semi-supervised topic modeling, sentiment analysis) to rigorously construct and evaluate our artifact. We have derived initial findings as contributions to theory, both kernel theory (e.g., Innovation Theory, Social Media Analytics Theory) and design theory. To contribute to a rather general and abstract knowledge base – "Nascent design theory" (Gregor and Hevner 2013) – and to design a purposeful artifact in a comprehensible way, we deduced Meta Requirements (MRs) and DRs (Gregor and Jones 2007; Heinrich and Schwabe 2014) for a trend analysis tool grounded in seminal works, resulting in the DPs.

Design and Development of the Artifact

Design Principles for a Trend Analysis Tool

The composition of MRs that describe "*what the system is for*" (Gregor and Jones 2007, p.325) is based on the purpose and scope of the trend analysis tool as discussed in the introduction. Thus, we define the solution objectives based on the problems of the investigation that our paper addresses. These MRs to be suitable for a class of artifacts are based on the current research literature (Gregor and Jones 2007; Heinrich and Schwabe 2014; Walls et al. 1992). Subsequently, we have synthesized DPs based on the DRs we have derived from the related work (i.e., extant literature) as well as based on DRs that are inspired by theory (cf. Möller et al. 2020; Purao et al. 2020). These DPs are defined as prescriptive statements, showing how to do something to achieve a certain goal (Gregor et al. 2020). The DPs we dispose fall into the category of action and materiality-oriented DPs that describe what an artifact should enable users to do and how the artifact should be built in order to do so (Chandra et al. 2015). For companies (= users) who are interested in conducting trend analysis based on unstructured social media data (= boundary conditions), and keeping our DRs for our artifact in mind, we derive six DPs for trend analysis tools:

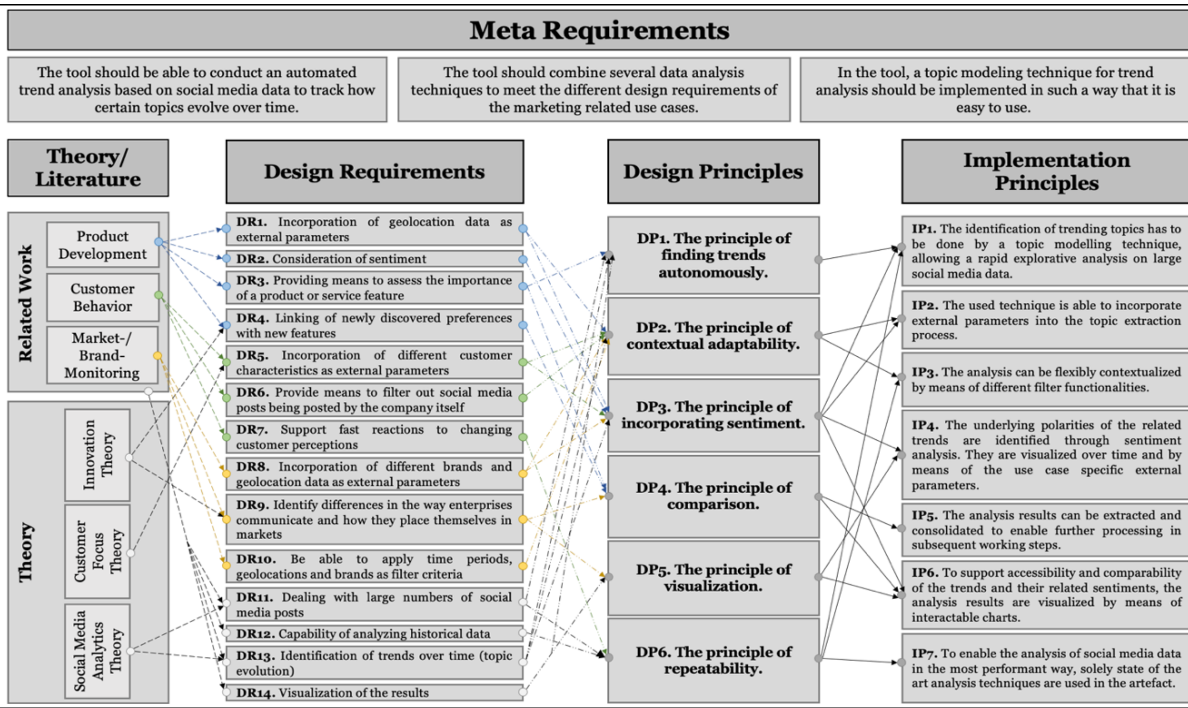

Figure 2. Design of an Automated Trend Analysis Tool for Unstructured Social Media Data

- **DP1: The principle of finding trends autonomously.** To track customers' evolving and changing requirements, it is necessary to retrace the evolution of topics. The tool should be able to find trends in large amounts of unstructured social media data and provide its users to apply autonomous trend analysis by including domain knowledge (i.a., seed words) to refine the identification of trending topics.
- **DP2: The principle of contextual adaptability.** Since external parameters (e.g., geolocations or customer characteristics) directly influence how a trend is pronounced, they must be included in the trend analysis tool. Therefore, the tool needs to provide the possibility to flexibly incorporate different external parameters and filters, so that the users can tailor the trend identification to their peculiarities.
- **DP3: The principle of incorporating sentiment.** To meet customers' expectations, it is necessary to identify issues evoking positive customer perceptions to retain them and to identify issues evoking negative customer perceptions to be improved. Thus, the tool needs to assign polarities to the individual trends and differentiate them over time and for the corresponding use case-specific parameters.
- **DP4: The principle of comparison.** To be competitive in the future, comparisons are necessary to identify the extent to which existing products and services meet customers' expectations (internal

comparison) and to compare a company's actions with those of competitors (external comparison). Thus, the tool should provide the retrieved analysis results to enable both kinds of comparisons.

- **DP5: The principle of visualization.** To obtain added value from the results and to benefit from the derived trends, they must be refined approachable. In doing so, the tool should be able to visualize the trends and sentiment values so that users are aided in interpreting and applying the derived results.
- **DP6: The principle of repeatability.** As a trend can change over time, the identification process should be executable often, and in a resource-saving way. Therefore, the tool should allow repetition at any time (including both historical and current data) to react quickly to changing circumstances.

These DPs are deduced from the DRs that are based on current research literature and theory. Gregor and Jones (2007) state that this reference to theory and literature discloses "*an explanation of why an artifact is constructed as it is and why it works*" (p.328). Based on related work and theory, we derive DRs our tool should meet. These DRs offer guidance by designing the artifact and advising the DPs (Böckle et al. 2021; Gregor and Jones 2007). These DPs refer to at least one DR and serve as an abstract "blueprint" of our artifact (Böckle et al. 2021; Gregor and Jones 2007; Heinrich and Schwabe 2014). By establishing these DPs, we made sure that they follow the value grounding (reference to the DRs) and the explanatory grounding (DPs are based on current literature and kernel theories) (Heinrich and Schwabe 2014). Furthermore, based on the instantiation of the DRs and DPs within our artifact MANTRA, we have defined Implementation Principles (IPs). In this way, we support "*the implementation in practice of an abstract, generic design method or development approach*" (Gregor and Jones 2007, p.329) in view of such an artifact for automated (topic modeling-based) trend analysis on unstructured social media data.

Technical Realization

To address the drawbacks of prevailing research and existing trend analysis tools, we have designed and developed the artifact *MANTRA* in the programming language *Python*. In the following, the instantiation of the DRs is described as they depict the required features of an automated topic modeling-based trend analysis tool in the most detailed way. Since the DPs and IPs depict a generic and prescriptive statement of how something should be done, they capture design-related knowledge and support the development of further IS (design) theories and new artifacts focusing on automated trend analysis. To reveal the association of the derived DPs/IPs to the DRs, their relations are pointed out as well.

Regarding the autonomous identification of trends **(DP1)**, the selection of the topic modeling technique to be applied is critical to success. As trends are time-dependent constructs, a rapid and explorative analysis must be provided. With this respect, unsupervised topic modeling techniques (e.g., LDA) are conceivable to enable an explorative analysis. However, given **DP2**, the incorporation of external parameters **(DR1, DR5, DR8)** is not feasible using unsupervised topic modeling. Beyond that, as stated by Chang et al. (2009) the potential of unsupervised topic modeling is stymied by their purely unsupervised nature, which often leads to topics that are neither meaningful nor effective at extrinsic tasks such as conducting a marketing campaign. Semi-supervised topic modeling techniques retain the flexibility of unsupervised techniques and can identify trends in an autonomous and performant way (Jagarlamudi et al. 2012). In doing so, they can incorporate external parameters such as geolocation information **(DR1, DR8)**, brands or products, and user-related external information **(DR5)**. As a semi-supervised topic modeling technique, *GuidedLDA* was chosen as it has achieved convincing analysis results within previous research (e.g., Ahadh et al. 2021; Chandrasekaran et al. 2020; Toubia et al. 2019). By providing seed information, such as a specific product/brand name or words related to an area such as the global food industry (e.g., burger sauce or burger patty), the topics generated by *GuidedLDA* will converge to gravitate into the contextual direction of those seed words, fostering a successful integration of identified trends into marketing decision-making.

Concerning use case (a) **Product Development**, the identification of trends and their matching with corresponding geolocations are required. Besides the incorporation of geolocation information through respective seed words, we, therefore, enabled MANTRA to pre-define different geolocations by applying a filter function, gaining maximum flexibility regarding the incorporation of external geolocation information **(DR1, DP2, IP3)**. This use case stipulates further the inclusion of sentiments within the analysis to determine the positive or negative tonality of the social media posts **(DR2, DP3, IP4)**. Therefore, we related to the "Valence Aware Dictionary for sEntiment Reasoning" (VADER) (Hutto and Gilbert 2014) technique. *VADER* is a lexicon- and rule-based sentiment analysis technique that is specifically attuned to sentiments expressed in social media and has achieved remarkable results (Hutto and Gilbert 2014). It can extract positive, neutral, and negative inflections in the range of -1 to +1. Additionally, the importance of a respective product or service feature has to be taken into account **(DR3, DP1, IP1)**. As probabilistic topic modeling techniques such as *GuidedLDA* infer their resulting topics based on various probabilistic distributions, depicting the relevance of the underlying topic words and thus the resulting topics (Crain et al. 2012), the assessment of the importance of the specific topic is met through the nature of topic modeling itself. Moreover, the linking of newly discovered preferences with new features **(DR4, DP4, IP5)** is compulsory. To develop an artifact that is suitable for various applications in marketing, including the processing of heterogeneous datasets, an automated linking of newly discovered preferences will not be feasible. Nevertheless, to take **DR4** into account, a function was implemented to extract the respective results into a Microsoft Excel file. In this way, the results are consolidated and can be used to manually map the identified preferences to new product or service features. Use case (b) **Customer Behavior Analysis** deals, i.a., with identifying behavior changes regarding different kinds of customers. Besides considering customer-related information by biasing the generated topics by means of seed words, users can pre-define customer groups based on attributes such as gender or age **(DR5, DP2, IP2)**. Further, a function to filter out posts created by the company itself to inhibit a bias in the analysis results **(DR6, DP2, IP3)** is required. Thus, MANTRA provides the possibility to dynamically define an identifier that is used to filter out associated data entries. **DR7 (DP6, IP7)** reflects a fast execution time when analyzing data concerning customer behavior. Here, MANTRA implements multi-threading, resulting in the parallel processing of independent tasks. Besides incorporating different brands and geolocation data with respective seed words **(DR8, DP2, IP2)**, (c) **Market-/Brand-Monitoring** requires identifying differences in the way companies communicate and place themselves within markets **(DR9, DP4, IP6)** by applying filter criteria such as time periods, geolocations, or specific brands **(DR10, DP2, IP3)**. Therefore, a filter function was implemented, handling the generation of different subsets of data concerning a specific time period or brand names. This allows comparisons based on diverse dimensions such as topics of interest or specific products and services. To enable the processing of large amounts of unstructured data **(DR11, DP6, IP7)**, only performant analysis techniques settled in the field of text mining are considered. Regarding the capability to view historical data **(DR12, DP6, IP7)**, flexible incorporation of different datasets including different time periods is feasible. The identification of trends over time **(DR13, DP2, IP2)** is considered by the possibility to determine specific time periods, resulting in a flexible identification of time-specific trends and their evolution over time. Finally, the trends, their use case-specific peculiarities, and all customizabilities are visualized **(DR14, DP5, IP6)** by a developed graphical user interface (GUI).

Demonstration and Evaluation of MANTRA

To assess MANTRA's capabilities, we followed our defined evaluation strategy *(see Steps "demonstration" and "evaluation" in Section "Research Procedure")*. In general, MANTRA is designed as an artifact to identify trends from various unstructured social media data. We decided to conduct the demonstration and evaluation on OCRs of the academic Yelp dataset, as it is a publicly available real-world dataset and has proven appropriate for verifying the capabilities of machine learning techniques (e.g., Binder et al. 2019; Heinrich et al. 2022). The academic Yelp dataset (Yelp 2021) comprises about 8.6 million OCRs of 160,585 businesses in different fields (i.a., restaurants, cultural sites, sports facilities) and contains different context information (i.a., creation time, federal state, author information) concerning these OCRs. Because this dataset has already been subject to trend analysis within the food and restaurant industry (e.g., Heinrich et al. 2022; Wörner et al. 2022), we decided to narrow our demonstration to this domain, resulting in a dataset of about 1.03 million OCRs. According to a 2018 survey by a panel of experts from the NRAB, *vegan cuisine* and *global cuisine* are likely to gain popularity in the US until 2030 (Statista 2019). By means of explorative analysis, we have next identified that restaurants with these two orientations and corresponding food offerings are actually included in the dataset. Together with the results from the previous trend analysis for the food and restaurant domain, this ensured that the possibility of MANTRA producing so-called false positives is considerably reduced (Pries-Heje et al. 2007). Consequently, MANTRA identifies topics for the two trends *vegan cuisine* and *global cuisine* because they are actually contained within the Yelp dataset.

The configuration *(see Figure 3)* represents the view when starting the tool and can be used to customize the trend analysis. The layout is aligned with the GUI design suggestions of Garrett (2010). *Section (i)* presents an excerpt of the dataset used, including the actual reviews and the accompanying business data (e.g., federal state or business name). To consider a representative analysis period concerning our demonstration, the analysis uses data from 2015 to 2021 *(ii)*, as the related NRAB survey conducted in 2018.

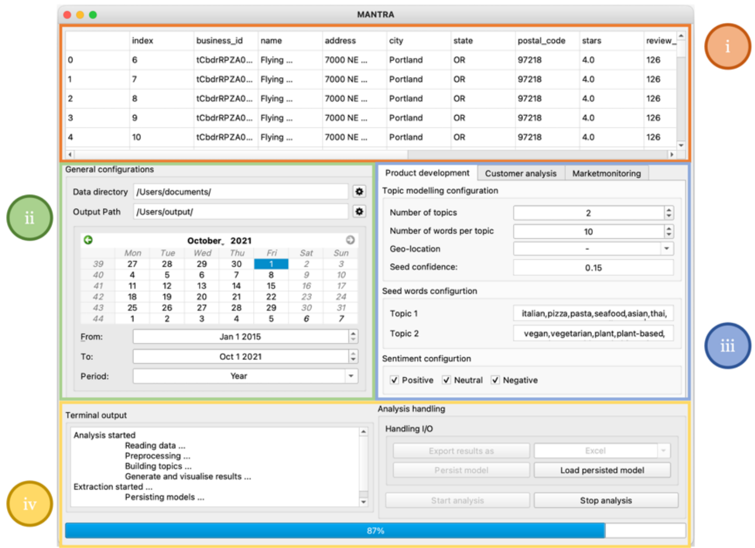

Figure 3. MANTRA – Configuration View

To incorporate the identified use cases, they were implemented in *section (iii)* using modular, dynamic tabs to enable a distinctive configuration. Here, regarding the use case (a) Product Development, the number of topics (*2*), words per topic (*10*), the respective external parameters of geolocation (*none*) and contextualized seed words (e.g., topic 1: Italian, Asian; topic 2: vegan, vegetarian) were defined. Moreover, all sentiment levels (*pos, neg, neu*) were considered. The start of the analysis, the handling of input and output (I/O) operations, and the monitoring take place in *section (iv)*. Once the process has been finished, the results are consolidated and displayed by responsive, interactive charts *(see Figures 4, 5, and 6)*.

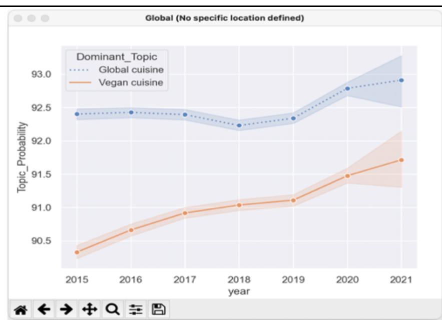

Figure 4. MANTRA – Identified Trends

Figure 5. Corresponding Sentiments (Vegan)

Figure 4 represents the determined topic probabilities with respect to the years, resulting in the tracking of the topic evolution based on their relevance. Here, both identified trends are continuously increasing in terms of their relevance. The trend *global cuisine* appears to slightly decrease from 2015 to 2018, but then continuously increases. The trend *vegan cuisine* increases steadily within the analyzed period. These results provide convincing evidence that both trends are on an upward trajectory and exhibit consistency with the findings described in the 2018 NRAB forecast. Moreover, the 2020 and 2022 Food & Health surveys by the International Food Information Council (IFIC 2020, 2022) show a significant percentage of US residents turned to a vegetarian/vegan diet relative to the previous years. As our tool detects a significant increase compared 2019 and 2020, convincing evidence is provided that our tool can identify even small deviations within the evolution of a trend. In addition, *Figure 5* comprises the trend-specific sentiments, by which a trend and the customer demands can be further deconstructed. All in all, our results confirm the general feasibility of MANTRA, as the elicited trends are consistent with the results of the two aforesaid surveys.

Trend topics	Products/Features (noun)	Descriptive characteristics (adj., verb)
vegan cuisine	food, restaurant, vegan, service, gluten, …	good, great, delicious, best, unfriendly, free, …

Table 3. Excerpt of the Resulting Excel

The analysis also comprises an Excel file to link the elicited results to new products/features. *Table 3* shows an excerpt of the resulting Excel file, containing the trend-related words and their classification into the part of speech for the topic *vegan cuisine*. This breaks down the trends into constituent parts, enabling the identification of specific products/features and related characteristics. As MANTRA depicts co-occurrences within the elicited trends by topic modeling, identifying sub-trends is feasible. Here, our analysis could, e.g., reveal that customers eating vegan also refer to gluten-free food in a vast frequency *(see Table 3)*.

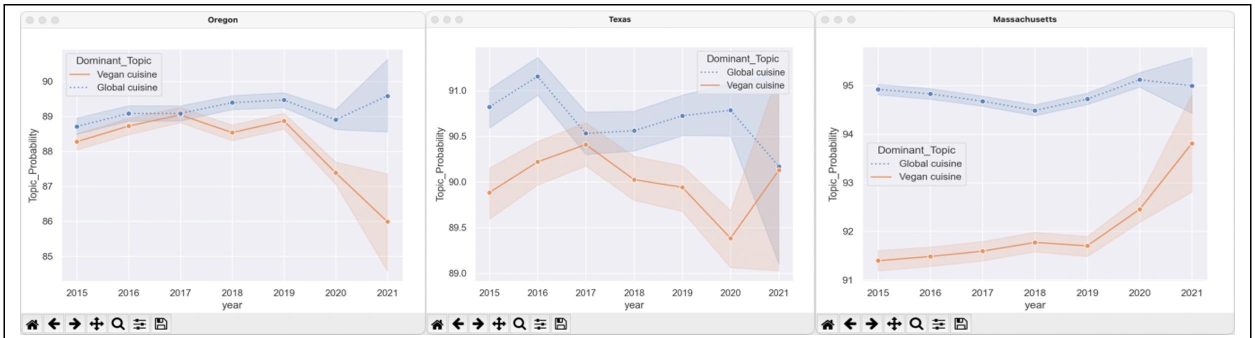

Figure 6. MANTRA - Identified Trends (From Left to Right: OR, TX, MA)

Since (a) **Product Development** refers to examining geolocation-based differences regarding the needs and expectations of customers, we further tested our artifact on sub-datasets referring to three different states in the US: *Massachusetts (MA), Texas (TX), and Oregon (OR)*. The results provide convincing evidence that MANTRA is able to identify geolocation-based discrepancies within the evolution of trends *(see Figure 6)*. It can be clearly seen that the topic *vegan cuisine* differs across the locations, as it consistently increases in MA but remains steady or even decreases in TX and OR. *Global cuisine* has declined since 2015 in TX, while slightly increasing in MA and OR. The sentiment distributions further reveal that – although the relevance of the vegan trend is vigorously increasing in TX from 2020 to 2021 – the respective reviews are negatively connotated *(pos: 0.27, neg: 0.61, neu: 0.12)*. In MA, the trend also increases sharply from 2019 to 2021 but is predominantly characterized by positive reviews *(pos: 0.56, neg: 0.3, neu: 0.14)*.

Generally, the development of our tool was based on the DRs, which are all met and technically realized *(see Figure 2 and the DPs and IPs)*. The 2018 and 2019/2020 trends could be verified by applying our tool to a representative real-world dataset, validating its functionality and therefore its practical applicability.

Discussion and Contributions

The implementation of the DRs and DPs has provided interesting results. As we have included the sentiment in our trend analysis tool, we could identify positive and negative inflections about the trends and how the sentiment values evolve over time. *Figure 5* shows that users spoke positively about the vegan trend in 2016. This positive resonance decreased until 2019, but in 2020 and 2021 the positive comments predominate again. Thus, our demonstration reveals that, although the vegan trend is consistently rising (in terms of topic probability) *(see Figure 4)*, the sentiment evolves distinctively. Without including the sentiment, one would not know that especially in 2019 ambivalent opinions were prevalent about the vegan trend and opening a vegan restaurant would not have yielded the success implicated by the upward trend.

To be successful and competitive in a targeted market, marketing representatives need to know which product and service features to design in which way to meet the global and local customers' expectations (Bae et al. 2018; Endo et al. 2015; Ha et al. 2017). This becomes particularly evident when comparing the three states of OR, TX, and MA *(see Figure 6)*. While both trends in MA have been growing steadily since 2015, the situation in TX and OR is different. In TX, the vegan trend declined between 2017 and 2020. However, in 2020, it started to gain importance. This is in contrast to the global cuisine trend which, peaked in 2016 but has dropped sharply since 2020. OR exhibits a completely different picture of the two trends. While the vegan trend initially remains stable and dropped sharply in 2019, the global cuisine trend remains stable over the years. The comparison of the three figures clearly indicates that flexibly including geolocation information at different levels of granularity (such as federal states) can have a considerable influence on whether and how to harness the trends. While aligning a restaurant to the vegan trend is (based on these analyses) advisable in MA, this is not the case in OR. To better assess these trends in TX and MA, we have combined the two alignments – geolocation and sentiment – because, as mentioned above, disclosing an upcoming trend does not directly mean that the trend exhibits necessarily positive connotations. Whereas the vegan trend in TX has gained in importance since 2020, the sentiment is predominantly characterized by negative contributions *(neg: 0.61)*. This indicates that either the vegan food offered so far does not meet the customers' expectations, or that vegan food is not well received in TX in general. In contrast to TX, the vegan trend in MA is predominantly characterized by positive reviews on average *(pos: 0.56)*. Without the inclusion of the geolocation data and the sentiment values, these aforementioned discrepancies in the results would not have been revealed.

Our investigation contributes to research and practice alike. As a contribution to practice, first companies can benefit from our comprehensive and modular artifact. MANTRA meets several DRs and, in contrast to the available solutions, does additionally provide a semi-supervised topic modeling technique *(GuidedLDA)*. Thus, MANTRA is capable of integrating several external parameters, which is deemed important in the extant research literature (Bae et al. 2018; Endo et al. 2015; Ha et al. 2017). Beyond that, MANTRA allows marketing representatives to integrate the use case-specific external parameters even at different granularity (e.g., federal states as geolocation) and therefore, i.a., to identify fine-grained discrepancies in customer perceptions. Moreover, we have implemented a functionality to link future customer preferences with new features (cf. DR4). By extracting the results of the trend analysis into an Excel file and classifying the trend-related words into the part of speech, trends and their constituent parts can be analyzed in more detail. For example, a company that has analyzed the trends that emerged in Section "*Demonstration and Evaluation of MANTRA*" could recognize that customers eating vegan also refer to gluten-free food with a vast frequency. As our results indicate, this functionality of a trend analysis tool is eminent as it supports companies, on the one hand, to identify fine-grained aspects of the elicited trends and, on the other hand, to support that the gained insights can be seamlessly included into the steps that are following the actual trend analysis (i.a., trend assessment and integration into the product and service portfolio of a company). In general, MANTRA has been developed in a modular way, enabling its users to define various settings through the corresponding GUI elements. For example, the time period for a specific trend can be set to conduct the analysis resource-efficiently. Several use case-specific DRs were implemented by offering modular and dynamic tabs to enable quick and almost effortless processing of common marketing-related use cases. As we have combined different machine learning approaches and designed our tool modularly, companies can adapt the analysis to their individual circumstances.

Furthermore, as an outcome of our DSR project, we achieved theoretical contributions to research that go beyond the technical contribution (i.e., the artifact). As we have designed a tool that is able to identify the voice of the customer, even about unknown customer demands and design features, our results also contribute to research theories such as Innovation Theory. Here, our approach can be included, e.g., in the stage-gate model of the innovation process that describes a conceptual and operational model for moving new product projects from idea to launch (Cooper 1996). Thereby, our results show that it is important to distinguish trends based on external parameters (e.g., geolocation), as customer perceptions may differ across global and local sites. By incorporating specific trend-related information within the different stages of the stage-gate model, the rigid specifications of stages and gates can be loosened up. Since social media enables customers to express their unfiltered opinions (Yan et al. 2014), thereon identified trends reflect customer perceptions, which can ensure that customers' needs are considered. Thus, by integrating the trend information in this way, the assessments at the go/kill checkpoints (i.e., gates) become less elaborate as the alignment with external requirements (e.g., customer wishes and demands) is maintained constantly. Moreover, Social Media Analytics Theory (especially concerning the research of automated trend analysis) states that the automated analysis of social media content still holds various challenges when it comes to its actual application (Fan and Gordon 2014; Stieglitz et al. 2018). With the design and development of our automated trend analysis tool, we purposefully combined various machine learning techniques (i.a., semi-supervised topic modeling, and sentiment analysis) and ensured that the DRs and the corresponding marketing-related use cases are technically realized. In this way, we contribute to Social Media Analytics Theory by proposing a need-fitted automated trend analysis for unstructured social media data.

The discussed results have shown the advantages of our tool "MANTRA" compared to existing ones, placing our DSR project in the group of improvements (development of new solutions for known problems) in terms of the DSR knowledge contribution framework of Gregor and Hevner (2013). DSR improvement projects contribute to both prescriptive theory, i.e., design theory (Gregor 2006), and descriptive theory, i.e., kernel theories such as Innovation- or Social Media Analytics Theory as described above (Gregor and Hevner 2013). By including GuidedLDA as a semi-supervised topic modeling technique within MANTRA that can flexibly integrate different external parameters as deemed crucial by our marketing-related use cases, our research as well affects the quadrant of exaptation (cf. Gregor and Hevner 2013). We have shown the proficiency of this technique and in this way contribute to IS and DSR research alike, as we propose semi-supervised topic modeling as a suitable technique for automated trend analysis (tools) to tailor the derived trends towards context factors and semantically towards the company and its fields of activity.

Based on the DRs derived from literature and kernel theories, DPs were formulated and proposed. By applying them during the design and development of the artifact followed by the demonstration and evaluation, an implicit empirical grounding of the DPs was achieved here (Heinrich and Schwabe 2014). Our DPs capture design-related knowledge and can thus support the development of further IS (design) theories and new artifacts. By considering e.g., DP2 Contextual adaptability, the importance of the contextualization (e.g., location) is highlighted. Since context directly impacts trend analysis results, the alignment with context leads to more meaningful insights. With the DPs, we made a first step towards contributing to design theory in terms of theory for design and action (Gregor 2006) as we comply with the conditions of March and Smith (1995) and Hevner et al. (2004), who pointed out under which a contribution to knowledge in DS has occurred: utility to a community of users, the novelty of the artifact and the persuasiveness of claims that it is effective.

Conclusion, Limitation, and Future Research

The early identification of new and auspicious ideas as well as trending topics regarding the development of products and services, the analysis of customers' behaviors, and the monitoring of markets and competing brands lead to competitive advantages for companies. However, prior literature and existing solutions for automated trend analysis on unstructured social media data do not incorporate semi-supervised topic modeling and do not sufficiently cover the different external use case-specific parameters as well as the specific requirements of the common marketing-related use cases (a) Product Development, (b) Customer Behavior Analysis, and (c) Market-/Brand-Monitoring.

Therefore, we identify several DRs and derive DPs. To technically realize the DPs, we combine several machine learning techniques and transfer these in an intuitive GUI to close the revealed research gaps *(see RQ1 and RQ2)*. Especially with the demonstration to about 1.03 million OCRs, we show that covering all DRs is essential for a target-oriented and feasible trend analysis. MANTRA also supports unsupervised trend analysis when no seed words are provided. Hence, also marketing representatives who have not gained prior knowledge of underlying trends can also conduct an explorative trend analysis using MANTRA.

Our investigation proposes contributions to practice and research alike *(see RQ2)*. Companies can benefit from our comprehensive and modular artifact, with which large amounts of unstructured social media data can be analyzed in a way best suited to the company's individual circumstances. We have also highlighted how our investigation made a first step towards contributing to design theory and kernel theories. Regarding Innovation Theory, the rigid sequence of stages and gates in the stage-gate model (Cooper 1996) can be loosened up by examining external requirements constantly. In this vein and according to the demonstration of MANTRA, it is exemplified that the stages should incorporate activities to assess the influences of the different external parameters (e.g., continents, phases of a day, personal characteristics; in the demonstration: federal states) on the perception of trends. By incorporating the DRs of a trend analysis tool, we enrich Social Media Analytics Theory with a need-fitted automated approach for analyzing social media content regarding the automated identification of trends.

It is worth mentioning, that the initial assessments on the practical applicability of MANTRA regarding (b) Customer Behavior Analysis and (c) Market-/Brand-Monitoring revealed promising results. Therefore, we will investigate these two use cases in more detail in the future and examine in more detail the potential of combining the three identified use cases with each other. In addition, to take the next steps toward a more mature design theory, we will first evaluate our artifact in a formative and artificial environment (e.g., a laboratory experiment). This allows us to improve our tool (whereby our DPs can be confirmed or adapted) before conducting a more elaborate evaluation in a more natural setting as a further part of the design cycle. There are also limitations to this research: Although we included a large set of investigations, we could identify probably even more use cases in further literature. Additionally, researchers from other fields could identify other use cases. Nevertheless, the identified use cases are undoubtedly important for marketing.

References

Ahadh, A., Binish, G., and Srinivasan, R. 2021. "Text mining of accident reports using semi-supervised keyword extraction and topic modeling," *Process Safety and Environmental Protection* (155).
Axtell, C., Holman, D., Unsworth, K., Wall, T., Waterson, P., and Harrington, E. 2000. "Shopfloor innovation: Facilitating the suggestion and implementation of ideas," *J. Occup. Psychol.* (73:3).
Bae, J., Havsol, J., Karpefors, M., Karlsson, A., and Mathiason, G. 2018. "Short Text Topic Modeling to Identify Trends on Wearable Bio-Sensors in Different Media Type," in *Proceedings of the 6th International Symposium on Computational and Business Intelligence*.
Bai, S., He, H., Han, C., Yang, M., Yu, D., Bi, X., and Panigrahi, P. 2023. "Exploring thematic influences on theme park visitors' satisfaction: An empirical study on Disneyland China," *J. Consum. Behav.* (22:2).
Batrinca, B., and Treleaven, P. 2015. "Social media analytics: a survey of techniques, tools and platforms," *AI & Society* (30), pp. 89-116.
Bhor, H., Koul, T., Malviya, R., and Mundra, K. 2018. "Digital Media Marketing Using Trend Analysis on Social Media," in *Proceedings of the 2nd International Conference on Inventive Systems and Control*.
Binder, M., Heinrich, B., Klier, M., Obermeier, A., and Schiller, A. 2019. "Explaining the stars: Aspect-based sentiment analysis of online customer reviews," in *Proceedings of the ECIS 2019*.
Böckle, M., Bick, M., and Novak, J. 2021. "Toward a Design Theory of User-Centered Score Mechanics for Gamified Competency Development," *Information Systems Management* (40:1), pp. 1-27.
Burmeister, K., Neef, A., Albert, B., and Glockner, H. 2002. "Zukunftsforschung und Unternehmen," *Praxis, Methoden, Perspektiven*. Essen.
Chandra, L., Seidel, S., and Gregor, S. 2015. "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions," in *Proceedings of the 48th Hawaii International Conference on System Sciences*.
Chandrasekaran, R., Mehta, V., Valkunde, T., and Moustakas, E. 2020. "Topics, trends, and sentiments of tweets about the COVID-19 pandemic," *Journal of Medical Internet Research* (22:10).
Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., and Blei, D. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models," *Advances in Neural Information Processing Systems*, pp. 288-296.
Chen, J., Yu, J., and Shen, Y. 2012. "Towards Topic Trend Prediction on a Topic Evolution Model with Social Connection," in *Proceedings of the International Conferences on WI-IAT*.
Cooper, R. 1996. "Overhauling the new product process," *Industrial Marketing Management* (25:6).
Crain, S., Zhou, K., Yang, S., and Zha, H. 2012. "Dimensionality Reduction and Topic Modeling: From Latent Semantic Indexing to Latent Dirichlet Allocation and Beyond," *Mining Text Data*, pp. 129-161.
Ding, K., Choo, W., Ng, K., Ng, S. 2020. "Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation," *International Journal of Hospitality Management* (91).
Duncan, T., and Moriarty, S. 1998. "A Communication-Based Marketing Model for Managing Relationships," *Journal of Marketing* (62:2), pp. 1-13.
Endo, Y., Toda, H., and Koike, Y. 2015. "What's Hot in the Theme: Query Dependent Emerging Topic Extraction from Social Streams," in *Proceedings of the 24th International Conference on the WWW*.
Fan, W., and Gordon, M. 2014. "The power of social media analytics," *Communications of the ACM* (57:6).
Garrett, J. 2010. "The Elements of User Experience: User-Centered Design for the Web and Beyond".
Gregor, S. 2006. "The nature of theory in information systems," *MIS Quarterly* (30:3), pp. 611-642.
Gregor, S., Chandra Kruse, L., and Seidel, S. 2020. "Research Perspectives: The Anatomy of a Design Principle," *Journal of the AIS* (21:6).
Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.
Gregor, S., and Jones, D. 2007. "The Anatomy of a Design Theory," *Journal of the AIS* (8:5).
Gulati, R., and Oldroyd, J. 2005. "The Quest for Customer Focus," *Harvard Business Review* (83:4).
Ha, T., Beijnon, B., Kim, S., Lee, S., and Kim, J. 2017. "Examining User Perceptions of Smartwatch through Dynamic Topic Modeling," *Telematics and Informatics* (34:7), pp. 1262-1273.
Heinrich, B., Hollnberger, T., Hopf, M., and Schiller, A. 2022. "Long-term Sequential and Temporal Dynamics in Online Consumer Ratings," in *Proceedings of the ECIS 2022*.
Heinrich, P., and Schwabe, G. 2014. "Communicating Nascent Design Theories on Innovative Information Systems through Multi-Grounded Design Principles," in *Proceedings of DESRIST 2014*.
Hevner, A., March, S., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105.

Hicks, A., Comp, S., Horovitz, J., Hovarter, M., Miki, M., and Bevan, J. 2012. "Why People Use Yelp.com: An Exploration of Uses and Gratifications," *Computers in Human Behavior* (28:6), pp. 2274-2279.
Hong, L., Ahmed, A., Gurumurthy, S., Smola, A., and Tsioutsiouliklis, K. 2012. "Discovering Geographical Topics in the Twitter Stream," in *Proceedings of the 21st international conference on WWW*.
Hong, L., and Davison, B. D. 2010. "Empirical Study of Topic Modeling in Twitter," in *Proceedings of the 1st Workshop on Social Media Analytics*.
Humphreys, A., and Wang, J.-H. R. 2018. "Automated text analysis for consumer research," *Journal of Consumer Research* (44:6), pp. 1274-1306.
Hutto, C., and Gilbert, E. 2014. "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *Proceedings of the International AAAI Conference on Web and Social Media*.
IFIC (2020). "2020 Food & Health Survey. 9th June 2020," *https://foodinsight.org/2020-food-and-health-survey/*. Lastly accessed on 6th April 2023.
IFIC (2022). "2022 Food & Health Survey. 18th May 2022," *https://foodinsight.org/2022-food-health-survey/*. Lastly accessed on 6th April 2023.
Irawan, M., Wijayanto, R., Shahab, M., Hidayat, N., and Rukmi, A. 2020. "Implementation of Social Media Mining for Decision Making in Product Planning Based on Topic Modeling and Sentiment Analysis," *Journal of Physics: Conference Series*.
Iwata, T., Watanabe, S., Yamada, T., and Ueda, N. 2009. "Topic tracking model for analyzing consumer purchase behavior," in *Proceedings of the 21st International Joint Conference on AI*.
Jagarlamudi, J., Daumé III, H., and Udupa, R. 2012. "Incorporating Lexical Priors into Topic Models," in *Proceedings of the 13th Conference of the European Chapter of the ACL*.
Jeong, B., Yoon, J., and Lee, J.-M. 2019. "Social Media Mining for Product Planning: A Product Opportunity Mining Approach Based on Topic Modeling and Sentiment Analysis," *Int. J. Inf. Manag.* (48).
Kaplan, A., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons* (53:1), pp. 59-68.
Khan, H. U., Nasir, S., Nasim, K., Shabbir, D., and Mahmood, A. 2021. "Twitter trends: a ranking algorithm analysis on real time data," *Expert Systems with Applications* (164).
Ko, N., Jeong, B., Choi, S., and Yoon, J. 2018. "Identifying product opportunities using social media mining: application of topic modeling and chance discovery theory," *IEEE Access* (6), pp. 1680-1693.
Lozano, M., Schreiber, J., and Brynielsson, J. 2017. "Tracking Geographical Locations Using a Geo-Aware Topic Model for Analyzing Social Media Data," *Decision Support Systems* (99), pp. 18-29.
March, S., and Smith, G. 1995. "Design and Natural Science Research on Information Technology," *Decision Support Systems* (15:4), pp. 251-266.
Mastrogiacomo, L., Barravecchia, F., Franceschini, F., and Marimon, F. 2021. "Mining quality determinants of product-service systems from user-generated contents," *Quality Engineering* (33:3), pp. 425-442.
Mayring, P. 2000. "Qualitative Content Analysis," *Forum Qualitative Social Research* (1:2).
Mayring, P. 2002. "Introduction to qualitative social research," *Weinheim: Beltz*.
Möller, F., Guggenberger, T. M., and Otto, B. 2020. "Towards a method for design principle development in information systems," in *Proceedings of DESRIST 2020*.
Moro, S., Lopes, R., Esmerado, J., and Botelho, M. 2020. "Service quality in airport hotel chains through the lens of online reviewers," *Journal of Retailing and Consumer Services* (56).
Mudambi, S. M., and Schuff, D. 2010. "Research note: What makes a helpful online review? A study of customer reviews on Amazon.com," *MIS Quarterly* (34:1), pp. 185-200.
Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. "Mine your own business: Market-structure surveillance through text mining," *Marketing Science* (31:3), pp. 521-543.
Park, E., Kwon, J., and Kim, S. 2021. "Green marketing strategies on online platforms: a mixed approach of experiment design and topic modeling," *Sustainability* (13:8).
Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3).
Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. 2012. "Design science research evaluation," in *Proceedings of DESRIST 2012*.
Pries-Heje, J., Baskerville, R., and Venable, J. 2007. "Soft design science research: Extending the boundaries of evaluation in design science research," in *Proceedings of DESRIST 2007*, pp. 18-38.
Puranam, D., Narayan, V., and Kadiyali, V. 2017. "The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors," *Marketing Science* (36:5).
Purao, S., Kruse, L. C., and Maedche, A. 2020. "The Origins of Design Principles: Where do… they all come from?, " in *Proceedings of DESRIST 2020*, pp. 183-194.

Qu, Q., Chen, C., Jensen, C., and Skovsgaard, A. 2015. "Space-Time Aware Behavioral Topic Modeling for Microblog Posts," *IEEE Data Engineering Bulletin* (38:2), pp. 58-67.
Reisenbichler, M., and Reutterer, T. 2019. "Topic modeling in marketing: recent advances and research opportunities," *Journal of Business Economics* (89), pp. 327-356.
Saha, A., and Sindhwani, V. 2012. "Learning Evolving and Emerging Topics in Social Media: A Dynamic NMF Approach with Temporal Regularization," in *Proceedings of the Conference on WSDM 2012*.
Sasaki, K., Yoshikawa, T., and Furuhashi, T. 2014. "Online Topic Model for Twitter Considering Dynamics of User Interests and Topic Trends," in *Proceedings of the Conference on Empirical Methods in NLP*.
Shi, L., Wu, Y., Liu, L., Sun, X., and Jiang, L. J. 2018. "Event Detection and Identification of Influential Spreaders in Social Media Data Streams," *Big Data Mining and Analytics* (1:1), pp. 34-46.
Sigala, M. 2012. "Social Networks and Customer Involvement in New Service Development (Nsd): The Case of www.Mystarbucksidea.com," *International Journal of Contemporary Hospitality* (24:7).
Sohn, B., Lim, H., and Choi, J. 2019. "The Prediction of Sales Volume and WoM Effect through Network Structure and Topic Modeling on Social Media," in *Proceedings of the PACIS 2019*.
Statista. 2019. "Ranking der prognostizierten Restaurant-Trends im Bereich 'Food & Menu'," *https://de.statista.com/statistik/daten/studie/1125284*. Lastly Accessed on 6th April 2023.
Stieglitz, S., Mirbabaie, M., Ross, B., and Neuberger, C. 2018. "Social Media Analytics – Challenges in Topic Discovery, Data Collection, and Data Preparation," *Int. J. Inf. Manag*. (39), pp. 156-168.
Sundbo, J. 1995. "Three paradigms in innovation theory," *Science and Public Policy* (22:6), pp. 399-410.
Tirunillai, S., and Tellis, G. 2014. "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research* (51:4).
Tseng, M., and Piller, F. 2003. "The customer centric enterprise," *Springer Science & Business Media*.
Toubia, O., Iyengar, G., Bunnell, R., and Lemaire, A. 2019. "Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by the psychology of media consumption," *Journal of Marketing Research* (56:1), pp. 18-36.
Tucker, C., and Kim, H. 2011. "Predicting Emerging Product Design Trend by Mining Publicly Available Customer Review Data," in *Proceedings of the 18th International Conference on Engineering Design*.
Ulbrich, F., Christensen, T., and Stankus, L. 2011. "Gender-Specific on-Line Shopping Preferences," *Electronic Commerce Research* (11:2), pp. 181-199.
Valdez, D., Padon, A., Barry, A., and Russell, A. 2018. "Alcohol Advertising on Twitter - a Topic Model," *American Journal of Health Education* (49:4), pp. 256-263.
Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: a framework for evaluation in design science research," *European Journal of Information Systems* (25:1), pp. 77-89.
Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research," *Communications of the AIS* (37:1), pp. 205-224.
Walls, J., Widmeyer, G., and El Sawy, O. 1992. "Building an Information System Design Theory for Vigilant EIS," *Information Systems Research* (3:1), pp. 36-59.
Wörner, J., Konadl, D., Schmid, I., and Leist, S. 2022. "Supporting Product Development by a Trend Analysis Tool Applying Aspect-Based Sentiment Detection," in *Proceedings of DESRIST 2022*.
Yan, X., Guo, J., Lan, Y., Xu, J., and Cheng, X. 2015. "A Probabilistic Model for Bursty Topic Discovery in Microblogs," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.
Yan, Z., Xing, M., Zhang, D., Ma, B., and Wang, T. 2014. "A Context-Dependent Sentiment Analysis of Online Product Reviews Based on Dependency Relationships," in *Proceedings of the ICIS 2014*.
Yang, Y., Yao, Q., and Qu, H. 2017. "Vistopic: A Visual Analytics System for Making Sense of Large Document Collections Using Hierarchical Topic Modeling," *Visual Informatics* (1:1), pp. 40-47.
Ye, H. J., and Kankanhalli, A. 2018. "User service innovation on mobile phone platforms: Investigating impacts of lead userness, toolkit support, and design autonomy," *MIS Quarterly* (42:1), pp. 165-188.
Yelp. 2021. "Yelp Open Dataset," *https://www.yelp.com/dataset*. Accessed in April 2022.
Zhang, H., Kim, G., and Xing, E. P. 2015. "Dynamic Topic Modeling for Monitoring Market Competition from Online Text and Image Data," in *Proceedings of the 21th KDD Conference*.
Zhang, P., Gu, H., Gartrell, M., Yang, D., Ding, X., Gu, N. 2016. "Group-based Latent Dirichlet Allocation (Group-LDA): Effective audience detection for books in online social media," *Knowl. Based Syst.* (105).
Zhao, K., Cong, G., Chin, J.-Y., and Wen, R. 2019. "Exploring Market Competition over Topics in Spatio-Temporal Document Collections," *The VDLB Journal* (28:1), pp. 123-145.
Zhong, N., and Schweidel, D. 2020. "Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model," *Marketing Science* (39:4), pp. 827-846.