NOVA IMS Information Management School

MDDM

Master's Degree Program in
Data-Driven Marketing

The Role of User-Generated Content in the Open Innovation Process.

Consumer-Centric Insights Driving Innovation.

Felipe Batitucci de Gusmão (m20211286)

Master Thesis

presented as a partial requirement for obtaining the Master's degree program in Data-Driven Marketing

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

The Role of User-Generated Content in the Open Innovation Process.

Consumer-Centric Insights Driving Innovation.

By

Felipe Batitucci de Gusmão

Master Thesis Work presented as a partial requirement for obtaining the Master's degree in Data-Driven Marketing, with a specialization in Digital Marketing and Analytics.

Supervised by

Diego Costa

NOV/2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Felipe Batitucci de Gusmão
Lisboa, November 30th, 2023

DEDICATION

To my beloved family,

This thesis would not have been possible without your love, support, and encouragement. I am deeply grateful for everything you have done for me and for the sacrifices you have made. Moving to Portugal was a big decision, and I appreciate how you all embraced the change with such resilience and positivity. Your willingness to adapt to a new culture, and to make new friends and connections, has made our journey all the more meaningful.

To my two children, Francisco and Cecilia, who have brought so much joy and laughter, thank you for your patience and understanding when I spent long hours on this thesis. I hope that someday you will look back on this period of our lives with fondness, knowing that we faced our challenges together as a family.

To my spouse, Luisa, rock and partner in everything, thank you for being my constant source of love and inspiration. Your unwavering support, understanding, and encouragement have sustained me through this journey.

Finally, I dedicate this thesis to all families who embark on a new adventure together, face life's uncertainties with courage and grace, and strive to create a better future for themselves and their loved ones.

With all my love,

Felipe Gusmão

ACKNOWLEDGMENTS

I want to take this opportunity to express my deepest gratitude and appreciation to all who have been instrumental in helping me decide on the subject of my thesis and supporting me throughout this journey.

First and foremost, I would like to thank my family again for their support, encouragement, and patience. Their love has been the foundation upon which I have built my academic, professional, and personal life. I am grateful to them for always being there for me and providing me with the necessary motivation and inspiration to pursue my dreams.

I would also like to thank my friends for their constant encouragement, support, and inspiration. They have been my pillars of strength, helping me overcome obstacles and challenges along the way. Their support has been invaluable in keeping me focused and motivated during the writing process.

Finally, I would like to acknowledge the support and encouragement of my work colleagues. Their guidance, feedback, and insights have been crucial in shaping my ideas and refining my research. I am grateful to them for providing me with the necessary resources and opportunities to pursue my research interests.

In conclusion, I am deeply indebted to my family, friends, and work colleagues for their support and encouragement throughout this journey. Without their help, this thesis would not have been possible. I am truly grateful for their presence in my life.

ABSTRACT

This master thesis examines the role of user-generated Content as a driver to feed the Open Innovation Process with relevant Consumer-Centric insights.

The study draws on a review of relevant literature, a data collection of 9600 reviews from Amazon Iberia, focusing on the Fast-Moving Consumer Goods (from hereon: FMCG) confectionery category, plus text mining analyses exercise identifying possible trends in product development respecting real consumer needs. An in-depth interview with five senior-level executives from different businesses that are known to have innovation as their core activities tries to figure out how important it is to have consumer opinion as the center of an Open Innovation strategy compared with the web scrapping results.

The findings suggest that User-Generated Content can be a powerful catalyst for Innovation. Still, its effective utilization requires careful consideration of the motivations and behaviors of users, the quality of data captured, the possible drawbacks of not having the proper context when using the information collated, as well as the Integration of appropriate incentives and governance mechanisms. The implications of these findings for companies are discussed and recommendations are provided for future research in this area.

KEYWORDS

Open Innovation; User-Generated Content; Social Media Marketing; Collaborative Innovation and Marketing Trends

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The concept of the Open Innovation model (from hereon: OI) was first mentioned in 2003 by Chesbrough as a new powerful methodology to share knowledge and collaboration, emphasizing the importance of inbound and outbound flows of information across the boundaries of the organization. These flows should guarantee access to external sources of cognizance, addressing the traditional innovation "out-of-the-box thinking" challenge (Osorno & Medrano, 2022). Leaders must understand that knowledge and expertise do not reside solely within the organization (Barrett & Tsekouras, 2022) and the collaboration process needs to be embraced as one of the most important stages when it comes to successful OI management (Yildirim et al., 2022) and that has a positive impact on Innovation, becoming an objective in itself (Audretsch & Belitski, 2022).

After two decades of studies and hundreds of academic papers, OI's interest has grown fast. Scholars running Ph.D. courses, conferences, and most top journals have an overwhelming appetite for OI special issues. However instead of covering diverse methodological branches and sources of information, most studies mainly focus on the benefits and risks of using OI.

At the same time, the world has witnessed an abnormal evolution in technologies such as Artificial intelligence (from hereon: AI) and Machine Learning (Dahlander et al., 2021), and the amount of data available through digitized text is growing in a fast-paced mode through online newspapers content, companies press releases, scientific articles as well as those created by users through ratings and reviews, comments in Social Media and so on, being available for research mostly free of charge (Antons et al., 2020).

Information system scholars found that digital technologies exist dynamically and are malleable, allowing data usage in many new ways to produce new products and services(Dahlander et al., 2021).

However, the vast majority of the studies in the OI field do not mention the importance of User-Generated Content (from hereon: UGC) as one of the cornerstones of the innovation process, even with companies aiming to have a consumer-centric approach (Ho-Dac, 2020).

Social Media shapes how companies consider insights from customers, users, and other stakeholders in the outside world (Dahlander et al., 2021). Consumer needs are a crucial part of marketing strategy, product development, and marketing research and should also be one of the main actors in the OI process. The astonishing evolution of online UGC can create the perfect storm for companies to identify new product opportunities and improve existing services and consumer needs using this data to specify attributes for conjoint analysis.(Artem Timoshenko et al., 2017).

A study on how UGC can affect the stages of OI is still lacking. At the same time, online consumer information plays a critical role in positively shaping future products and services

and moving towards closing this gap between the two main streams in the OI paradigm (Ho-Dac, 2020).

This study aims to understand better the value of having UGC in the OI process, taking into account the methods to extract this information from the big lake of online data to develop new products and services.

The study begins with a systematic search and overview of the most recent literature on OI, UGC, and Text Mining over the past four years (from 2019 to 2023) with a multi-step selection process to identify and analyze the most relevant academic study from top journals (ABS3/4).

Secondly, the paper shows a case study through Text Mining, identifying what consumers are talking about products in FMCG, specifically confectionery, using 9600 Ratings and Reviews from Amazon Iberia—understanding the similarities and the gap between the Consumer's needs and the Company's roadmap. There is no intention to have a powerful insight but just to validate the methodology.

It will also count on qualitative research using an in-depth interview with five Senior Leadership Members of different businesses exclusively responsible for both Brand Marketing and/or Innovation departments to understand their decision-making process regarding Innovation and OI and if they either use or consider the usage of UGC and how it affects the products/services development considering all stages of the process.

This study contributes to the OI literature and methodology, adding a new element to the Innovation process in companies, truly considering consumer needs; we end this study by paving the way for future research, showcasing the substantial lack of information about this topic, and proposing new branches of knowledge studies.

This paper is structured in five sections using a mixed method: First, we went through a systematic literature review focused on the recent scholar's studies from 2019 to 2023, considering more than 40 selected academic articles, always from top journals (ABS3/4), mainly with subjects related to OI, UGC, and Text Mining.

This theoretical work aims to seek the key elements of each topic and understand the perks and risks of using UGC data in the OI process and the similarities and incongruencies between them. The second part considers a case study based on the Text Mining process identifying insights about the product food business, collecting 9600 Ratings and Reviews from Amazon Iberia regarding the same topic.

Then, in-depth research with top decision-makers aims to understand their point of view better, how they are open to adopting a new element into the innovation process, and how they see the potential risks and benefits of putting consumer needs at the center of their business strategies, answering the following questions: RQ1: What is your current methodology when driving innovation? RQ2: How can UGC be used to improve the OI

process considering consumer-centric insights? RQ3: In which OI process phase should the UGC be considered? RQ4: How open are the senior-level executives from top Companies to using UGC in their innovative process? RQ5: How can UGC help to put the Consumer needs at the center of the strategy?

Next, the study tries to correlate both results and propose if and where the UGC should be implemented in the OI process.

Finally, we present our conclusions and encourage future researcher opportunities to use UGC within the OI process.

2. LITERATURE REVIEW

2.1. OPEN INNOVATION

According to (Zhang et al., 2023), in their study 2023, learning is divided into exploitative, an in-depth investigation of existing knowledge, and exploratory, absorption, and transformation of new knowledge in new ideas. The first involves selecting, implementing, improving, and refining existing knowledge. The second one is responsible for seeking, discovering, creating new company knowledge, and trying new opportunities. Improving knowledge has encouraged enterprises to seek information from OI activities.

The innovation process is a crucial component of a competitive edge between companies. For a long time, it was supported by internal department studies, relying on the capacity to conduct research to develop new products and services. But in the past twenty years, collaborative work has raised a new way to innovate (Ogink et al., 2022). The evolution of technologies and the increase in available data revolutionized how companies acquire information internally and by exchanging it with external actors. This paradigm, called OI, enables value proposition creation through these new collaborative working methods (X. Zhang et al., 2023). The formalization of this collaborative process is key when it comes to OI management (Yildirim et al., 2022).

Proposed by Chesbrough in 2003, the OI model was created to reinforce the relevance of the inbound and outbound flow of information across the organization's borders during an innovation process. It can bring many benefits, such as accelerating time-to-market, enriching know-how, and potentially reducing risk with cost-sharing (Puliga et al., 2022). In a recent review of this concept, he defines OI as follows:

"Open Innovation is the purposeful usage of inbound, and outbound knowledge flows to accelerate internal Innovation and to expand the trade markets for external innovation use. This paradigm assumes that organizations may and must use external and internal ideas, as well as inner and outer paths to trade, in their pursuit to shift their technology" (Osorno & Medrano, 2022, page 439).

Many companies already have based most of their knowledge on OI, like Intel which has collaborated with a global open innovation platform where, with universities, they explore alternative technological applications, acquire external knowledge, validate existing ones, and explore new business opportunities, that can later be integrated into their business model (Osorno & Medrano, 2022).

According to a report from the world's largest crowdsourcing platform, Eyeka, 85% of the best 100 global brands have used UGC as one of their sources of inspiration to generate innovative

product/service ideas. Companies like Lego, Starbucks, Xiomi, and DELL maintain a crowdsourcing community just to get better insights from the user (H. Zhang et al., 2022).

But the process of open Innovation can be tricky, and depending on the approach to follow, companies can be unable to deliver the Innovation at the right time or even not meet customers' expectations, like the Boeing 787 Dreamliner case where the Company decided to work with a high number of outside suppliers as well as with internal R&D team and overtime the project suffered with delays and structural flaws, creating a cost increase, and risk for the innovation novelty (Madanaguli et al., 2022).

In this case, governance can play an important role in ensuring that the open innovation process is managed to maximize the benefits for all stakeholders involved. In a high-level approach, governance can stand for licensing, acquisitions, R&D contracts, spinouts, and corporate venture capital. Still, with the increase of new ways of doing OI, it can be expanded, but not limited to, to hackathons, innovation contests, crowdsourcing, and social networks (Cavallo et al., 2022).

OI is an interactive process where actors exchange resources to create value; called a co-creation process and has been identified as a critical part of enhancing Innovation. However, for this process to work at one size, we must have companies facilitating inbound and outbound flows with a clear governance model. On the other hand, a full network must exist that enables actors to iterate and collaborate toward a common goal, gradually displacing the traditional competitive model(Osorno & Medrano, 2022).

The potential usage of UGC to enable external knowledge search and exchange spans all stages of the innovation process, from ideation to the final solution (Barlatier et al., 2022). And with the benefit of not having communication between the Company and the consumers, facilitating the governance of the information collected (Ho-Dac, 2020).

Understanding the stages of the OI process is important to determine the collaboration type to be used. Defining the process flow (inbound OI, outbound OI, and coupled OI) can help the Company benefit from the knowledge they don't have internally, reducing the time spent and costs and increasing the Company's know-how. The collaborative stage lies at the heart of the OI process. It can be done through either scientific-based collaboration (R&D, Scientific Organizations, Universities, Public and Private Research centers, etc.) or market-oriented (such as Suppliers, Costumers, Consumers, and even Competitors) (Yildirim et al., 2022). Collaboration is "the process through which two or more actors engage in a constructive management of differences to define common problems and develop joint solutions based on provisional agreements that may coexist with disagreement and dissent" (Audretsch & Belitski, 2022).

But even with a well-structured OI process in place, it can hide pitfalls like the motivation of all actors involved, which can significantly increase the project costs, intellectual property management, information structure, governance across many different stakeholders, quality

assurance, trust and reliability of the information and many others associated risks (Osorno & Medrano, 2022).

Risks can also arise due to the lack of control and protection policies when using data from this massive amount of collected information from the diverse actors in the OI process. The well-accepted DART process (Dialogue, Access, Risk Assessment, and Transparency) can summarize it, outlining a new way of promoting organizational co-creation (Madanaguli et al., 2022).

Digital collaboration has grown along with the fast-paced evolution of emerging digital channels (e.g., Social Media, Digital Platforms, and Services), but besides having a huge amount of data that can be captured through digital channels and help the OI process, this astonishing information can also reveal digital traps because it can be cumbersome or even incorrect. The lack of analytics capabilities can also cause companies to have the right information but not be capable of extracting relevant insights (Li et al., 2022).

As this study aims to cover part of the OI process based on UGC data collection, it is important to highlight the risks identified related to data-driven decisions. Data risk lies in how data is collected from actors in the OI process and, most importantly, how they are used inside the organization. One study related to data risk (Madanaguli et al., 2022) has divided it into three types of menace: Data Privacy Risk, Data Distortion, and technical risks (Madanaguli et al., 2022).

The limitation of this study lies in Data distortion that occurs when data is poorly translated from the collection point to the phase where this data is used as input in the OI process. The lack of reliability can be part of the discrepancy between what was expected vs. what was extracted from the sources where UGC was collected. One way to address this problem is to allow anonymous participation where personal data and opinions are not considered when analyzing the data. This reinforces the importance of using UCG as part of the OI process (Madanaguli et al., 2022).

Nowadays, a huge amount of data is created daily in Social Networks. Some corporate sites and new relevant publishers are taking advantage of this massive information, known as UGC, to feed their platforms with pertinent information that can drive users to their web properties—considering that as a useful resource to identify insights and create knowledge through evidence collected from Social Media mining. For that reason is imperative to define the role that UGC plays in the OI process and how OI can benefit from such a powerful source of information(Saura et al., 2022a).

2.2. User-Generated Content

One of the most known concepts for Social Media is the highly interactive web platforms through which communities and individuals can collaborate, share, co-creating, discuss, and

modify UGC and dynamic exchange knowledge between individuals and organizations(Barlatier et al., 2022).

In a highly connected digital environment where you can find an almost unlimited amount of data generated on a daily basis, Social Networks have been seen as a perfect storm to explore the world, create knowledge, and extract insights on both traditional and trendy topics(Saura et al., 2022b).

Scholars also argue that UGC can increase technological knowledge through Data Mining (Saura et al., 2022b), impact innovation capabilities, and improve standards of practice (Barlatier et al., 2022).

UGC content aggregated from a collaborative network provides useful information about people's attitudes, opinions, and behaviors, accelerating knowledge sharing and opinion formation, triggering collective wisdom, and speeding up the OI framework's decision-making process (Hsiang & Rayz, 2022).

UGC is also a way for people to have online relevance, like a social credential. Based on Social Network Analyses, a Social Network comprises nodes and ties representing the type and level of connection between actors. Moreover, Social Network Analyses tell us that the actor's position in the social network influences the relevance and importance of their ideas in a group of people. It is also real when talking about crowdsourcing and idea generation, which the network centrality concept can capture. People with a high degree of centrality can be expected to be more influential (H. Zhang et al., 2022).

Some studies highlight the importance of using UGC instead of Firm Generated Content throughout the marketing consumer funnel, positively impacting the awareness, consideration, and purchase intent phases (Colicev et al., 2019).

But also Social Media is also a term aggregating many different platforms, including Social Network sites, blogs and microblogs, forums, professional Social Networks, collaborative sites, and sharing sites. (Barlatier et al., 2022). This study will be based on Amazon.es 9600 Ratings and Reviews in the confectionary category to understand how a specific category works and demonstrate consumers' opinions that can help improve a product-based development on consumer needs. Also, social scientists have a good consensus that the Content generated in Social Media can provide new solutions for old problems or even create a connection between available knowledge and academic contributions (Saura et al., 2022b).

In order to create a well-accepted product and have meaningful insights using both business intelligence and marketing analytics process, a database of online consumer behaviors have started to be considered a confident source of information with consumer content coming up from publishing Content, expressed opinions, and requested information feeding a huge lake of data (Saura et al., 2021)(Hsiang & Rayz, 2022).

Social publishers focus on exchanging information and typically cover one-to-many communications and new possibilities of a one-on-one approach with personalized Content (Barlatier et al., 2022). On the other hand, social communities support knowledge exchange and co-creation, where knowledge is recognized as more meaningful information that can also create value and has raised the importance of shaping and spreading public information (Hsiang & Rayz, 2022).

While we can see an astonishing number of successful cases of collaborative creation, a long-term analysis finds that relying on this type of information over time is very challenging as most of these ideas generated by creators tend to repeat (Ho-Dac, 2020). Recent studies show that companies are taking advantage of mining information from these sources (Saura et al., 2022b), obtaining knowledge at a very low cost and putting the Consumer at the center of their strategy, involving them in the innovation process(Brunetti et al., 2020) and adding the Consumer in the development phase by creating complex patterns of knowledge transfer between innovator seekers and providers across different stages of the innovation process (Barlatier et al., 2022). While the lake of information created by UGC could seem messy and unstructured, a bunch of data mining techniques based on Machine Learning have been applied to identify key elements that can evolve the OI process (Saura et al., 2021).

Several studies affirm being impossible to create knowledge about the perception and utility of a new product without involving customers' points of view (Barlatier et al., 2022). Consumers' needs have been an important part of marketing strategy(Artem Timoshenko et al., 2017). The difference between using UGC is that users do not send any requests or solutions to the companies. Instead, they are just sharing their opinions with others on Social Media, so companies just need to listen to their conversation and extract and organize the most useful data (Ho-Dac, 2020).

Online consumers tend to be very passionate about discussing issues they are interested in. The information generated by them can also be useful for contributing to product development by giving insightful ideas and spreading and discussing the latest trends (Hsiang & Rayz, 2022).

Consumers' needs are traditionally achieved through interviews and focus groups or heuristics like managerial judgment or product comparison review. It means that a classical approach goes through multiple experiential interviews or focus groups. Human analyses review the transcript to identify needs, remove redundancies, and finally, identify a hierarchical structure for consumers' needs. Still, recently, the explosion of UGC has created both demand and opportunities to complement those old approaches to analyzing previously unavailable data collected online. Among many advantages, the usage of UGC has many benefits. Online Content is easy and cheap to collect. It also includes competitors' information as well as information related to broader and/or narrowed categories (Artem Timoshenko et al., 2017).

UGC should also be used for refining ideas about product development through online and on-time feedback, being one of the possible game changers in product innovation, converting UGC into an innovative resource, which can be a competitive edge against market competitors as Companies believe that crowd wisdom can foresee the future market for their products. (Lin et al., 2022)

2.3. UGC TEXT MINING

Unstructured data in digital Content is rapidly increasing in volume and relevance in research on Innovation. Text mining presents a set of solutions to help innovators explore a large scale of data efficiently. Somehow it combines qualitative and quantitative data by structuring, analyzing, and understanding textual data on a large scale (Antons et al., 2020).

Users are no longer empowered to post their reviews and ideas as virtual interaction grows. Still, they are also involved in product evaluation that can benefit companies through two-way communication highlighting users' responsiveness, interactivity, and engagement (H. Zhang et al., 2022).

With new technologies arising, the data-driven business model will play a more important role in business development and Innovation in companies. The exponential data increase requires more technology when treating UGC to use this type of information in the OI framework (Dahlander et al., 2021).

Natural language processing, deep learning, and computational science have been well-developed and widely adopted and can be used to analyze and mining UGC from different sources of Social Media, to reduce the manual process workload, helping to efficiently and cost-effectively convert UGC into Innovation, and serving as a useful tool when collecting relevant insights according to companies needs. According to the organizational learning theory, a company's ability to absorb new knowledge and put it to use - known as absorptive capacity - is key to driving Innovation. This includes the potential to acquire and assimilate further information and the ability to transform and exploit it (Li et al., 2022).

Developments in Natural Language Processing and mainly unsupervised topic modeling techniques can potentially solve some of these manual limitations. This technique is also recognized in many areas of social science as a powerful way to analyze large amounts of unclassified data (Lu & Chesbrough, 2022).

Idea mining uses computational methods to automatically extract novel and innovative ideas from unstructured text. In the Internet environment, there are two main types of idea mining (Gupta et al., 2019). The first involves identifying creative ideas from various UGCs on the web, including online product reviews and tweets. The second type of idea mining involves analyzing ideas' Content, structure, and category to enable enterprises to utilize them for product innovation. The ultimate goal of both kinds of idea mining is to convert large volumes

of internet data into valuable innovation resources for businesses, which can be achieved using natural language processing techniques, machine learning techniques, and ontologies (Lin et al., 2022).

Analyses of data from the Danish Company Lego also show how Machine Learning and Text Mining can detect new ideas in online communities to be used in the OI process using past and current data into a predictive model (Dahlander et al., 2021).

Using the Machine Learning technique for Text Mining unstructured data can help discover latent topics, scan a set of digital environments, detect word and phrase patterns, and automatically cluster word groups and similar contents that best characterize a bunch of information. And it can be useful when discovering user needs to develop new products into the OI process. There are several ways to mine data from the web, including various techniques and algorithms. One method that has become increasingly popular is the Latent Dirichlet Allocation method. This is because it can handle large-scale data and identify latent topics, which can be interpreted and analyzed. The Latent Dirichlet Allocation method has been recognized for its benefits in research and innovative processes (Lu & Chesbrough, 2022).

Applying Text Mining to explore variables and improve the measurement of existing variables can be interpreted as a signal of significant maturity of the mode and methodologically accepted to collect UGC and analyze the data accordingly, generating useful insight to complement the OI framework in the toolbox of the innovative process (Antons et al., 2020).

In Figure 1, the study shows an approach for Text Mining that will be used in this article to process information from Amazon related to the confectionary category, collect useful data about product innovation, and compare it with the results of in-depth research with top decision-makers in Innovation. In this structure, we showcase the process phases of data gathering, data processing, content analyses, and Integration of the mining results and the results of this study.

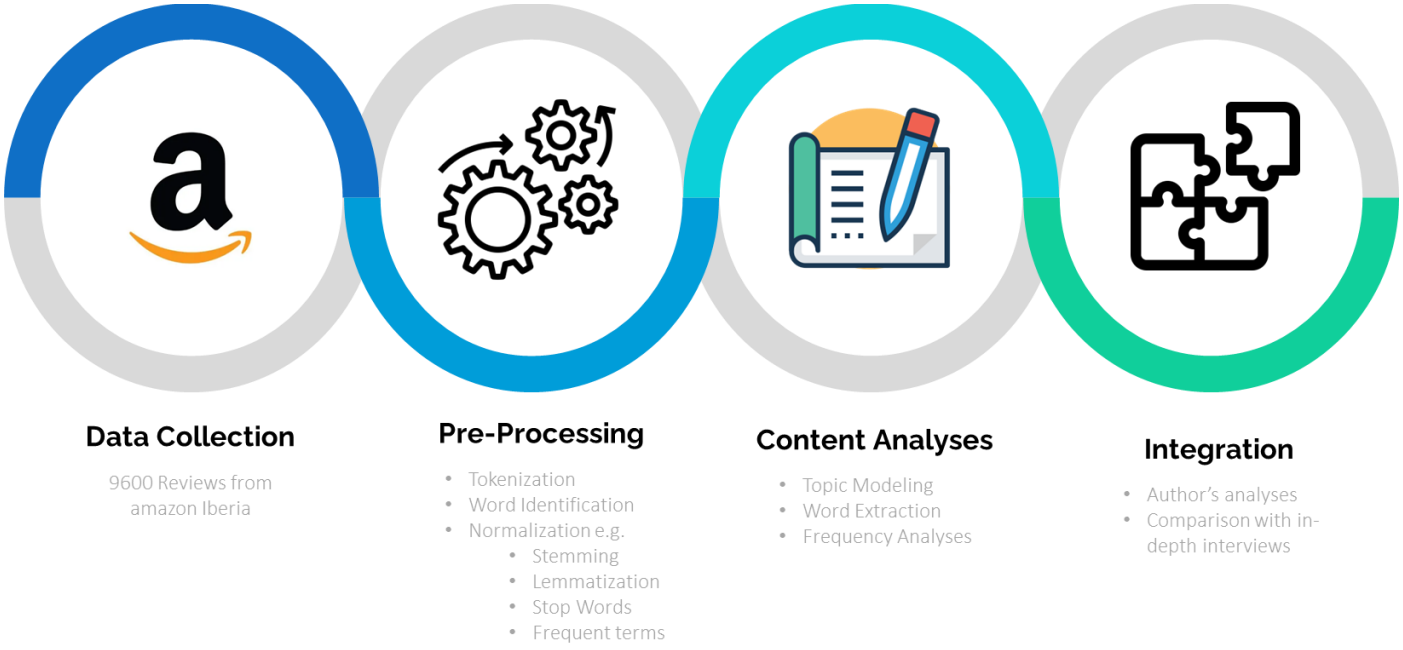

Figure 1. TM Process, technique, and Integration

3. METHODOLOGY & FIRST RESULTS

3.1. GUIDING RESEARCH

In this early stage of the study, we applied a Systematic Literature Review methodology to find, catalog, and analyze papers from top journals (ABS3/4), funneling down the understanding of how OI, UGC, and Text Mining can work together to step up the OI process.

The Systematic Literature Review methodology also provided a systematic overview of the papers, guiding the next steps of this study. Based on the same approach found in the paper Diving into the Uncertainties of Open Innovation: A systematic review of Risks to uncover pertinent typologies and unexplored horizons (Madanaguli et al., 2022), the study ran through 5 steps, (1) developing a guiding research question to be answered, (2) identifying keywords, (3) defining what should be included or removed, (4) identifying relevant papers and finally (5) reading and analyzing one by one to drive better how OI can benefit from UGC in its Innovative process (Mengist et al., 2020).

3.2. GUIDING RESEARCH QUESTIONS

The review of top journal papers (ABS3/4), covering topics related to consumer needs, UGC, OI, and Text Mining, intends to find a clear correlation between these topics, paving the way to understanding better how to collect useful information from the web, with low cost and not depending on governance. The study intends to prove that UGC usage in the OI process can benefit the business, putting the Consumer at the center of the strategy. Moreover, the study relies on the part of this Systematic Literature Review to answer some Research Questions: RQ1: What is your current methodology when driving innovation? RQ2: How can UGC be used to improve the OI process considering consumer-centric insights? RQ3: In which OI process phase should the UGC be considered? RQ4: How open are the senior-level executives from different businesses to using UGC in their innovative process? RQ5: How can UGC help to put the Consumer needs at the center of the strategy?

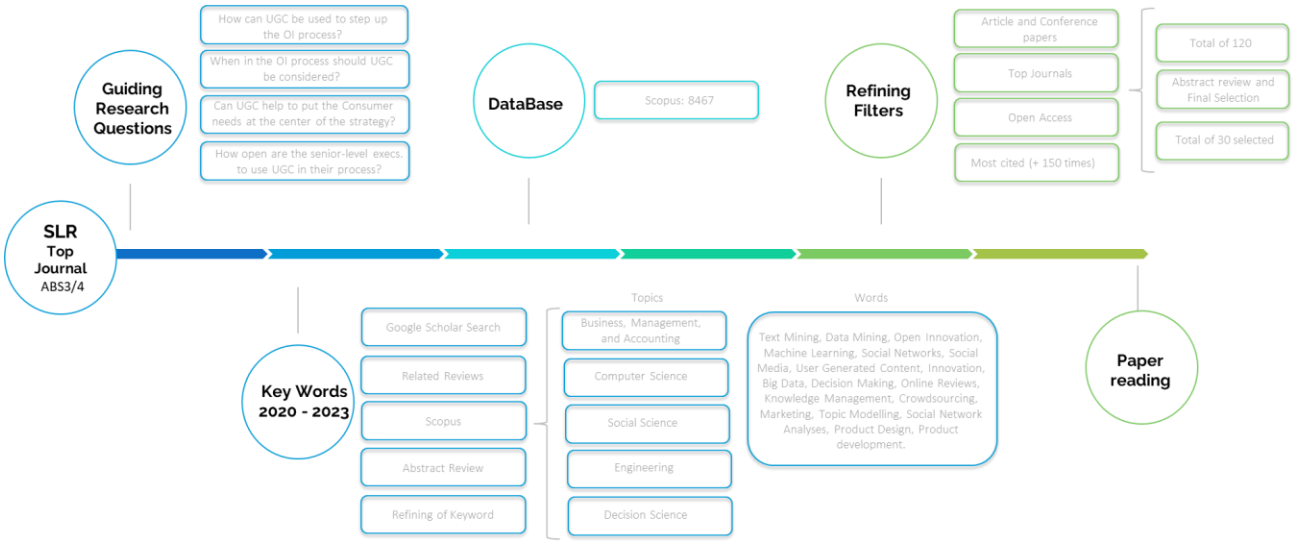

Figure 2. SLR Methodology.

3.2.1. KEYWORDS

After defining the question that would drive the study, we used Google Scholar (https://scholar.google.com/) and Scopus (https://www.scopus.com/search/form.uri?display=basic#basic) to select the main criteria to be used, having a first look in Related Reviews and Abstract reviews. Based on the recommendation mentioned in one of the papers to structure the SLR (Madanaguli et al., 2022), this first outlook used keywords like Open Innovation, User Generated Content, Crowdsource, Collaborative, Product development, Machine Learning, and Text Mining. Then, we used similar words, such as UGC, Consumer needs, etc., to increase the robustness of data collection as recommended by (Mengist et al., 2020). We found more than 14.000 papers related to the Research Questions from this process.

Trying to narrow down the huge number of papers, we restricted documents from 2020 to 2023 and decided to focus on Scopus to find relevant articles to be analyzed. We also first filtered in Scope to define five main topics and selected Keywords that drove the study to 8467 pertinent articles on this topic

3.2.2. FILTER, INCLUSION, EXCLUSION, AND RELEVANT PAPERS

Based on those 8467 papers found in Scope and considering the focus on one platform, we were able to go deeper into the analyses to synthesize our intention to link UGC to OI by examining validated research. Therefore we set some criteria to refine filters and get more driven information for the study: (1) We selected just Conference Papers and Articles, as one of the intentions was to use valid references to support the theoretical background of this study., (2) Using Academic Journal Guide (from hereon: AJG) (https://charteredabs.org/academic-journal-guide-2021/), we double check through the

title, if the selected document was part of the top journal following the ABS3/4 criteria to guarantee that the study relies on most valuable information, (3) selection of just Open Access documents, to speed up the analyses without dependence of paid terms, (4) selection of top-cited documents considering as a threshold paper with more than 150 citations which niched our study in just 120 articles being more manageable for review, and (5) A deep dive in these 120 abstracts to define the top 30 most related documents to be used in this study.

3.2.3. PAPERS READING AND FIRST ANALYSES

As a result of this Systematic Literature Review methodology, we found four main subjects that corroborate the importance of using UGC in the OI process and that need to be analyzed through a case study using Text Mining and in-depth research with senior-level executives from different businesses: (1) Relevant studies about UGC and its impact in the consumer funnel, (2) The importance of having OI in the product/service development, and the risks and benefits of relying upon external data and the necessity of a clear Governance Model, (3) The relevance of having consumer needs as the center of the business strategy and (4) How companies are using Text Mining and Machine Learning to extract useful information and not only use this information back in Social Media but also extract relevant insights that can be used as a game changer in their Business Model creating a clear competitive edge in their sectors. The Literature Review is a combination of different points of view that can guide us in the topic of this study and also can open the case for future research, according to new technology evolution and trend analysis like the usage of ChatGPT and Non-Fungible Token, in the OI Framework, for instance.

3.3. TEXT-MINING

Text Mining, mentioned in the Literature Review, is a powerful technique to extract valuable insights and knowledge from large collections of unstructured or semi-structured text data. With the exponential growth of Digital Content, businesses, and organizations have access to vast text data that can provide valuable information about Customer Behavior, Market Trends, and other crucial facts.

Text Mining techniques involve using algorithms and statistical models to analyze and extract patterns, themes, and relationships from data, which can help businesses make data-driven decisions and improve their overall performance. From Topic Modeling to Frequency Analysis, Text Mining techniques have become essential in various fields, including marketing, finance, healthcare, and Social Media Analysis.

This study considers the composition of two complementary techniques, a mixed method, Data Collection through Web Scraping and Text Mining, to identify topics that could be included in the OI process. It does not intend to identify a specific insight as an output. Still,

it verifies if, through an automated analysis taking Consumers Rating and Reviews as an example, we can collect useful information in a well-structured way that can shed light on the usage of Consumer insights benefiting the whole innovation chain.

3.4. Data Collection

For this analysis, the study utilized a code in Python, using Anaconda Navigator with Jupyter Notebook, and an API connector to extract 9600 Consumers Ratings and Reviews from Amazon Iberia using Amazon's SKU number (ASIN) from 2016 to 2023. This amount was defined by narrowing down the options using the following criteria: (1) FMCG products, (2) Confectionary Category, (3) Chocolate, (4) Chocolate Bars, and (5) Big Brands: Cadbury, Crunch, Ferrero, Galak, Kinder, Kit Kat, Lindt, Lion, M&M, Maltesers, Mars, Milka, Milkbar, Nesquik, Nestle, Snickers, Toblerone, and Twix as following:

ASIN	Page	ProdTitle	ProdLink	AvgRating	TotalRatings	TotalReviews	RevTitle	RevBody
B0762ND2BZ	1	Cadbury Dairy Milk F	https://www.amazor	4.7	18	4	Buenisimo	Me encanta soy adicta
B0762ND2BZ	1	Cadbury Dairy Milk F	https://www.amazor	4.7	18	4	El Mejor chocolat	El mejor chocolate que
B0762ND2BZ	1	Cadbury Dairy Milk F	https://www.amazor	4.7	18	4	Delicioso chocola	Rápido envío. Chocolat
B0762ND2BZ	1	Cadbury Dairy Milk F	https://www.amazor	4.7	18	4	Attention, titre c	Le produit considéré c
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Calidad Supremo	Si te gusta el chocolate
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Por fin una forma	Me ha encantado, no e
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Muy bueno	Es un chocolate muy du
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Chocolate.	Muy buen producto.
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	The best chocolat	Best bar of chocolate th
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Chocolate	Chocolate was nice but
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Great but didn't r	Great stuff but unfortu
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Arrived deformec	Arrived, melted and de
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Arrived in poor cc	Arrived in a very poor c
B00NGKWI42	1	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Good chocolate	I gave this to my brothe
B00NGKWI42	2	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	good	good as a gift
B00NGKWI42	2	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Fab fab fab	Simply yummy
B00NGKWI42	2	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	How can I choc ba	$21 for one choc bar
B00NGKWI42	2	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Cadbury's chocola	I bought theses as stoc
B00NGKWI42	2	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Chocolate gods	cadburys, need i say m
B00NGKWI42	2	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Four Stars	value for money
B00NGKWI42	2	Cadbury Dairy Milk,	https://www.amazor	4.6	528	18	Five Stars	Thank you xxxxx

Table 1. Sample of Extraction from Amazon Iberia.

Before starting the Text Mining process, all the Spanish comments were translated from Spanish to English using the GOOGLE TRANSLATOR feature in GOOGLE Sheets to guarantee data homogeneity in 80% of the comments and guarantee a fair amount of data to be treated by English-based Python algorithms. Other languages were not translated.

Then, using the model in Figure 3 below, the analyses went through a Topic Modeling Analysis, Word Extraction, and Frequency Analysis to identify relevant Content that could be included in an OI process, putting Consumers' information at the heart of the OI process.

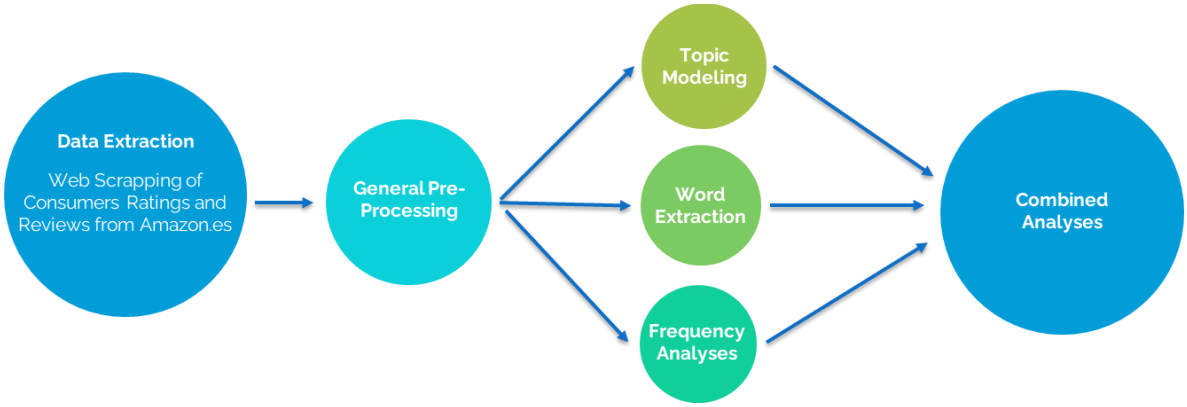

Figure 3. Extraction to Analyse Model.

4. FIRST RESULTS

4.1. TOPIC MODELING

Before diving into the Topic Modeling Analyses, the study verified the collected Data quality throughout two Analyses: Perplexity and Coherence.

The methodology involved an iterative process, where it continually refined the selection of topics based on the Perplexity and Coherence of the resulting clusters. By evaluating the Perplexity and Coherence of the topics, it is possible to identify and eliminate any issues arising from noisy or irrelevant data.

The approach proved effective in analyzing text data and identifying patterns in the language used in customer reviews. Through this process, it is possible to gain valuable insights into customers' experiences and identify areas where improvements could be made to enhance customer satisfaction.

Perplexity is an intrinsic evaluation metric frequently used for language models. It measures how surprised the model is when encountering new data it has not previously seen. This measure is calculated as the normalized log-likelihood of a test set held out.

Regarding the log-likelihood aspect, perplexity can be viewed as a measure of how likely new, unseen data is given the previously learned model. In other words, it evaluates how well the model represents the statistics of the held-out data. There is no target for perplexity but studies have shown that a lower perplexity score indicates better generalization performance. The perplexity of this study after running the model was approximately -8.3, as shown in Figure 4.

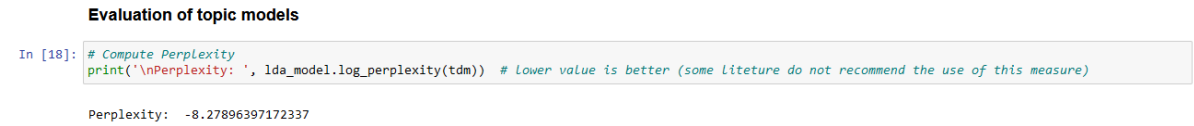

Figure 4: Perplexity score of 9600 Ratings and Reviews from Amazon Iberia.

Another way to evaluate the quality of the topics produced by a Topic Modeling method is to use Coherence measures. Coherence measures aim to capture the degree of semantic similarity between the words within a topic. The coherence value measures the semantic coherence of the words within a single topic by using the cosine metric to calculate their similarity, which ranges from 0 to 1. The topic coherence score measures the degree of semantic similarity between high-scoring words in the topic.

Coherence measures help to distinguish between semantically interpretable topics and those that are artifacts of statistical inference. The coherence score can measure how interpretable the topics are to humans, with topics represented as the top N words with the

highest probability of belonging to that particular topic. The coherence score depends on the data used to calculate it, and there is no one way to determine whether a score is good or bad. Generally, we want to maximize the coherence score.

The coherence score usually increases with the number of topics, but the increase becomes smaller as the number of topics gets higher. The elbow technique can achieve the trade-off between the number of topics and the coherence score. The coherence score also depends on the Linear Discriminant Analysis (LDA) hyperparameters, such as alpha, beta, and gamma. Machine learning hyperparameter tuning techniques can be used to optimize the coherence score, but manual validation is always necessary.

The coherence concept combines multiple papers into one framework that evaluates the coherence of topics inferred by a topic model. Despite its limitations, coherence is a valuable tool for assessing the interpretability of topics generated by topic models. The coherence score, as shown in Figure 5 below, after running the data set was 0.4 which can be considered a good rate, having as a base the range between 0 and 1 above-mentioned.

```
In [19]: # Compute Coherence Score
         coherence_model_lda = CoherenceModel(model=lda_model, texts=corpus, dictionary=id2word, coherence='c_v')
         coherence_lda = coherence_model_lda.get_coherence()
         print('\nCoherence Score: ', coherence_lda)

         Coherence Score:  0.4362286443949433
```

Figure 5: Coherence score of 9600 Ratings and Reviews from Amazon Iberia.

When cross-referencing the coherence score and the number of topics found in the data analyses, and also considering the elbow in Figure 6 below, the higher the number of topics, the better, this study focused on using eight topics to identify the main relevant subjects for the consumers.

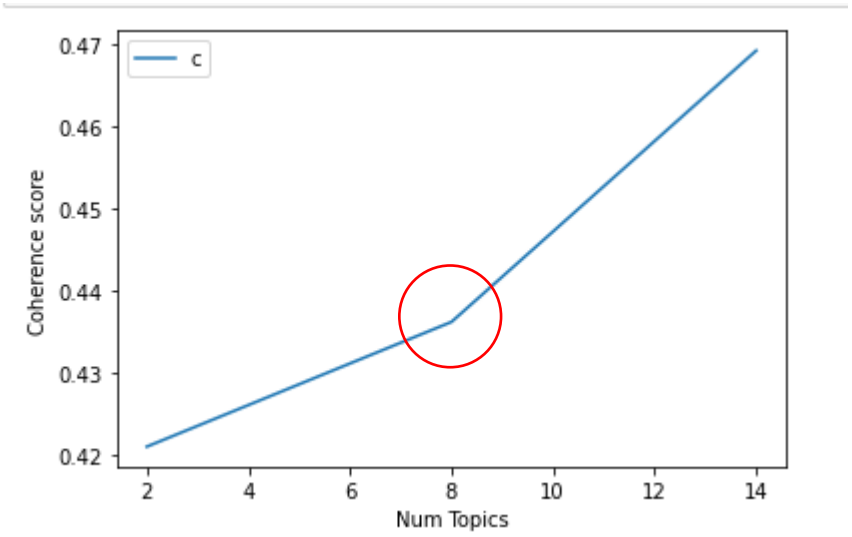

Figure 6: Matrix Coherence Score vs. Number of Topics to be analyzed.

The next stage of this study, after analyzing the quality of the data collected, was to run a preprocessing methodology on the 9600 Ratings and Reviews, cleaning the extracted information by removing through an English-based Python code, punctuation, special characters, HTML, numbers, and line breaks, and afterward, identifying cluster words for a given set of Content.

Then, using the Linear Discriminant Analysis (LDA) method, which involves the formulation of hypotheses about similar topics based on the usage of similar words, the code mapped each review in our extract file to issues encompassing many of the document's words.

Tables 2 and 3 below show the first result in aggregating all the information on eight main topics, considering the ten main words that belong to each topic and their weight within the respective topic (e.g., in Topic 1, the word product has 0,017 of importance among the other nine words within the same topic).

Topic 1	Topic 2	Topic 3	Topic 4
0, '0.037*"il" + 0.026*"è" + 0.026*"non" + 0.020*"che" + 0.017*"prodotto" + 0.017*"money" + 0.017*"original" + 0.014*"la" + 0.014*"sono" + 0.013*"packaging	1, '0.041*"quality" + 0.027*"excellent" + 0.025*"bad" + 0.023*"et" + 0.022*"je" + 0.018*"piece" + 0.017*"pour" + 0.016*"got" + 0.015*"bon" + 0.014*"très"'	2, '0.041*"die" + 0.029*"ich" + 0.029*"und" + 0.022*"lecker" + 0.022*"ist" + 0.019*"da" + 0.018*"der" + 0.017*"nicht" + 0.017*"schokolade" + 0.015*"sie"'	3, '0.049*"super" + 0.047*"little" + 0.039*"come" + 0.039*"always" + 0.038*"per" + 0.037*"e" + 0.036*"also" + 0.035*"milka" + 0.034*"white" + 0.025*"amazon"'

Table 2. First Topic Modeling Analyses considering the first four topics.

Topic 5	Topic 6	Topic 7	Topic 8
4, '0.084*"chocolate" + 0.046*"good" + 0.031*"bar" + 0.022*"like" + 0.020*"taste" + 0.017*"love" + 0.016*"buy" + 0.013*"time" + 0.012*"well" + 0.012*"n\'t"'	5, '0.063*"price" + 0.055*"one" + 0.038*"le" + 0.030*"best" + 0.029*"bag" + 0.026*"lindt" + 0.025*"get" + 0.024*"eat" + 0.021*"toblerone" + 0.018*"de"'	6, '0.026*"christmas" + 0.024*"£" + 0.021*"cocoa" + 0.019*"black" + 0.019*"loved" + 0.019*"go" + 0.017*"open" + 0.015*"think" + 0.014*"high" + 0.013*"star"'	7, '0.052*"great" + 0.050*"product" + 0.037*"box" + 0.028*"perfect" + 0.028*"di" + 0.024*"give" + 0.024*"value" + 0.023*"gift" + 0.020*"size" + 0.014*"arrived"

Table 3. First Topic Modeling Analyses considering the last four topics.

Considering the purpose of this study and the analyses mentioned above, a first glance, we can see that topics related to **Quality**, **Price**, **Delivery**, **Taste**, and **Seasonality**, like Christmas, should be considered when brainstorming for the next product innovation in the confectionary category. But the topic Modeling analyses can give us more granularity information.

By selecting the Top 30 most Salient terms within each topic and checking the estimated term frequency within the chosen topic, we can verify the relevance of words on topics one, three, four, six, and seven, the ones that give us more information about Consumer's needs and can help us to be focused in the stages of OI process.

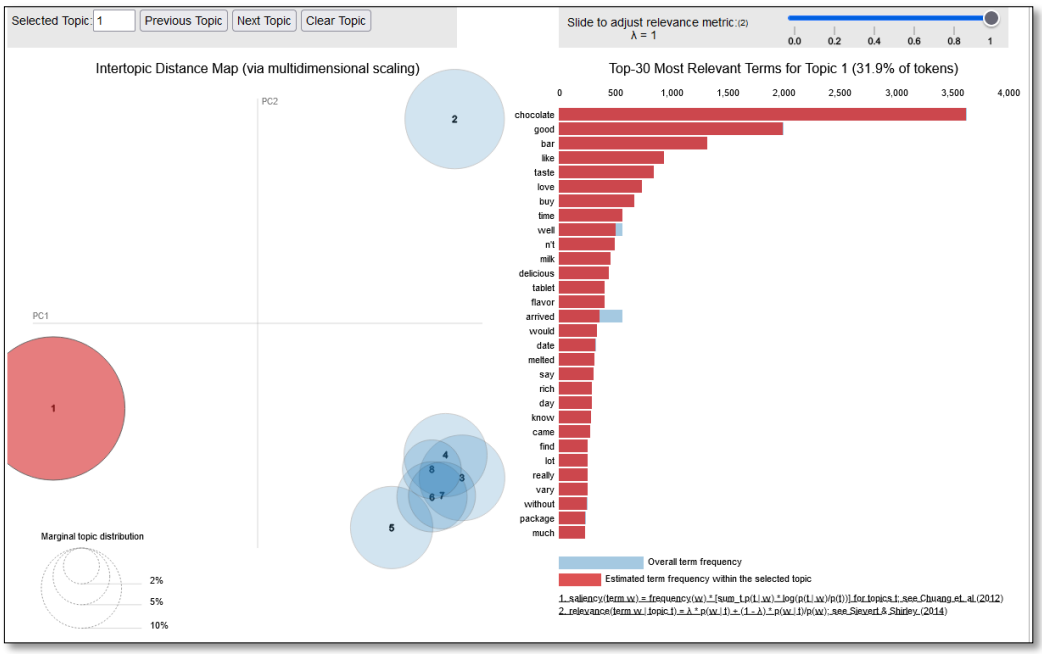

Figure 7: Frequency within Topic 1

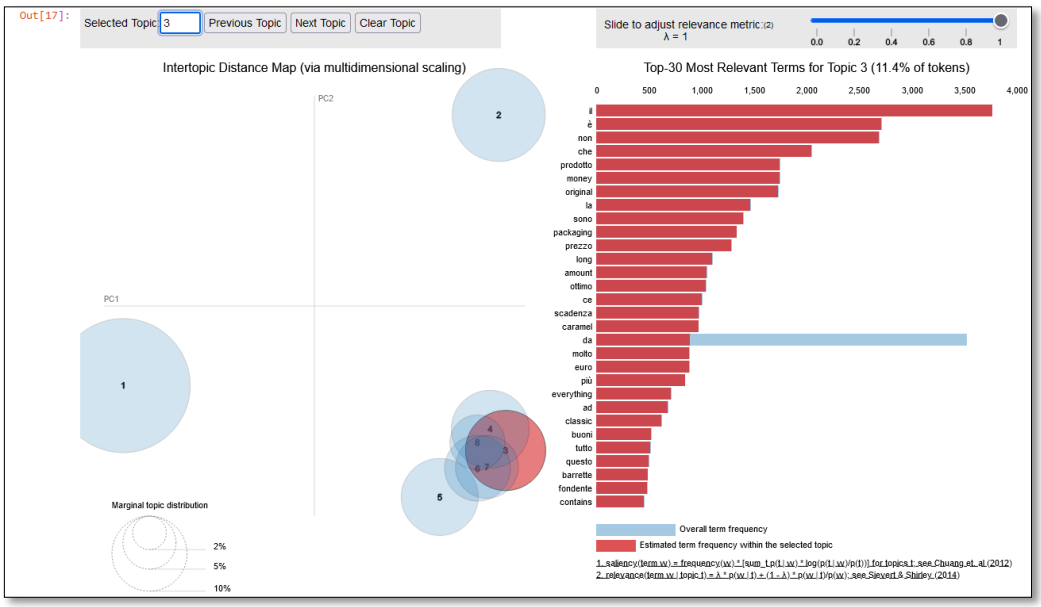

Figure 8: Frequency within Topic 3

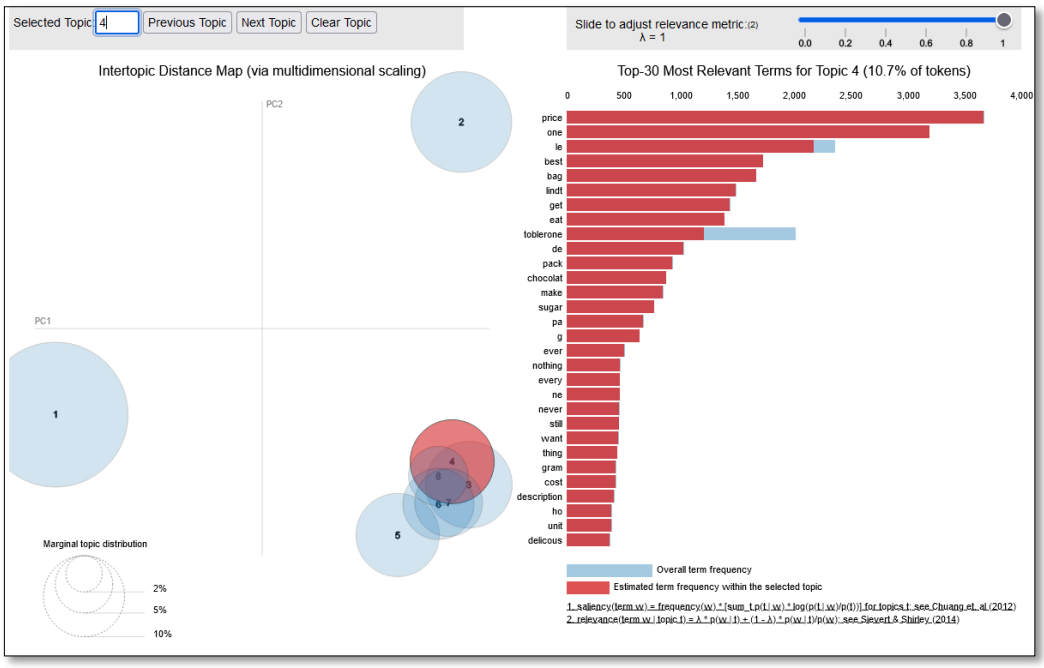

Figure 9: Frequency within Topic 4

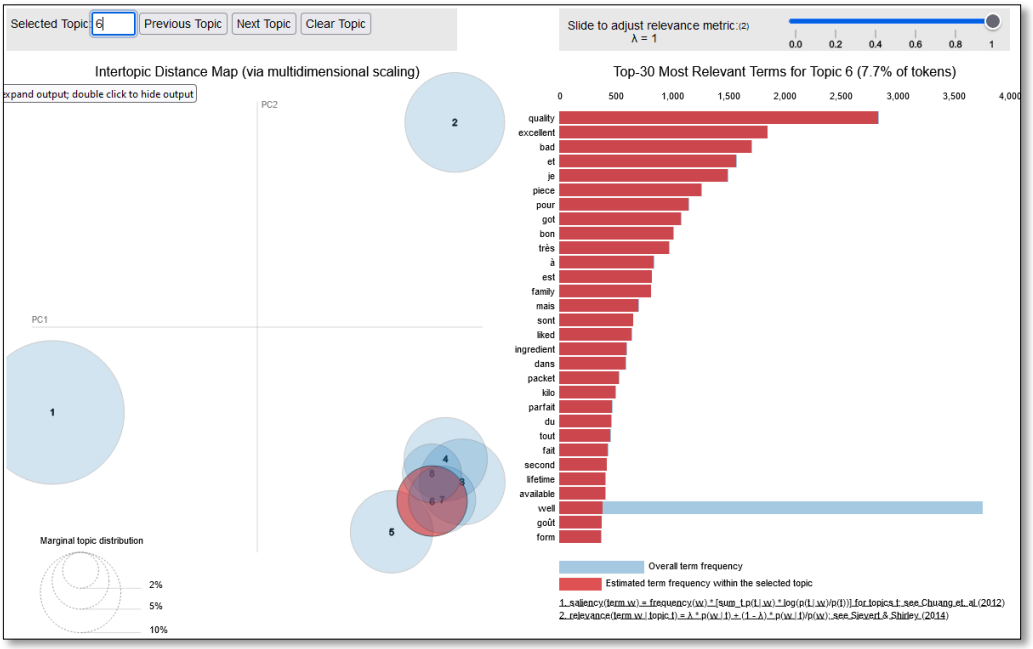

Figure 10: Frequency within Topic 6

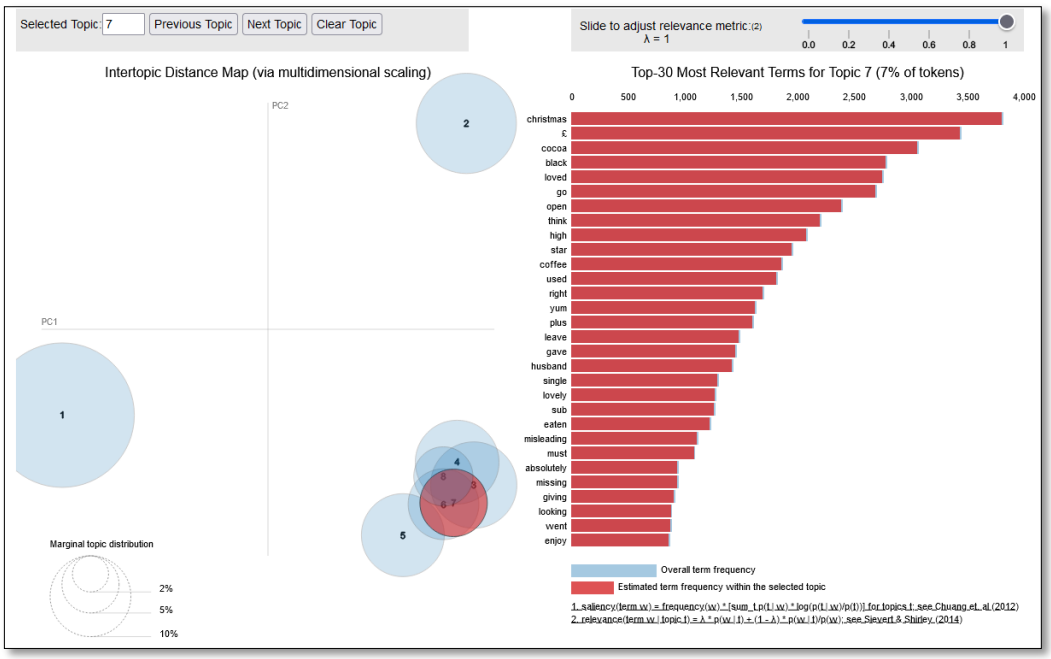

Figure 11: Frequency within Topic 7

4.2. KEYWORD EXTRACTION

Keyword extraction is a simple but effective technique that defines specific words to be identified in a large data pool, understanding their context better. There are many methodologies to run Keyword Extraction, but to perform a Keyword analysis; this study used the RAKE method, which stands for Rapid Automatic Keyword Extraction. Considering the previously identified as relevant for this study, the terms Product, Innovation, Package, Flavor, New, Good, Bad, Expensive, Cheap, and Price were selected to compose the RAKE model as per the example in Figure 12 below.

```
In [13]: # The first 10 reviews
         # termsToSearch = ['product', 'innovation', 'package', 'taste', 'flavor', 'new', 'good', 'bad', 'expensive', 'cheap', 'price
         ppText_searched[0:10]
```

```
['fast shipping . extraordinary chocolate of a prestige brand',
 'if you like milk chocolate ... cadbury is the brand that has the
best quality for me . it is a pity that in peninsula it is only
achieved through amazon and in some supermarket , but it is very good
.',
 'very good product .',
 'good as a gift',
 'the chocolates are very good i love twirl because it is like puff
pastry , the only bad thing is that discards in a thousand pieces
arrived .',
 'the pack came not as advertised , it came with a price written on it
of 1.25 pounds , so i didn t get what i have orded .',
 'world renewned cadburys chocolate at a budget price . the twirl
sticks are not reduced in size eithher . go for it !',
 "cadbury have done it again , these twirl bars have shrunk to the
size of lark `` twiglets ''",
 "would not recommended ordering all broken do n't taist good and took
ages to come",
 'the taste itself is absolutely lovely . my mum is from the uk and i
used to get used time we visit my grandmother . so for my birthday , i
went ahead to treat myself to sub sweet childhood memories . i love
cadburys and i do love the flakes & twirls . i must be honest , when i
ordered , i wrongly assumed i wouled be getting the packs with two
fingers . my fault entirely but the price is vary steep for what you
get . so beware , it is only 5 slithers that you will be receiving .']
```

Figure 12: Words identified in the first ten reviews.

Considering the insights gathered from 9600 Reviews, this more qualitative analysis corroborates the proposals of using the topics mentioned above to be included in the analyses of the OI process.

4.3. Frequency Analyses

Lastly, adding to the qualitative analyses part, the study sorted the words in descending order to understand the most frequently mentioned terms in the Amazon Iberia Rating and Reviews according to the criteria above-mentioned.

Table 4 clearly shows terms like Good, Like, Taste, Price, and so on as the most frequent among the 9600 Ratings and Reviews collected.

Term	Frequency		
chocolate	3549	bars	634
good	1789	product	633
like	956	lindt	619
taste	843	und	596
tablet	803	di	564
great	729	il	547
die	710	buy	544
love	689	cocoa	540
price	687	ich	539
one	674	flavor	528

Table 4: Frequency Analysis - Top 20 Most Frequent Results

The same result can have the relevance and frequency of the words better visualized through a Wordcloud, Figure 13, considering all the most frequent words that appeared in the collected reviews.

Figure 13: WordCloud for top 20 results

Yet, it is also possible to draw relevant insights by analyzing the most frequent terms in reviews, more specifically in the combination of two and three words (Bi-grams and Tri-

grams). By exploring Bi-grams and Tri-grams, we can have more texture in the most relevant aspects mentioned in Reviews.

In Bi-grams, Figure 14, considering the top 50 terms encountered, the consumers' insight to be used in the OI process starts to get clear, with phrases like "good price," "value money," "good quality" and "taste good" being highlighted with high frequency in Ratings and Reviews.

```
# PLOT top 20 most frequently occuring BI-GRAMS
top_words = get_top_n_words(processedReviews['PreProcessedText'], n=50, nGrams=2)
top_df = pd.DataFrame(top_words)
top_df.columns=["Word", "Freq"]
sns.set(rc={'figure.figsize':(10,8)})
g = sns.barplot(data=top_df, x="Freq", y="Word", color="grey")
```

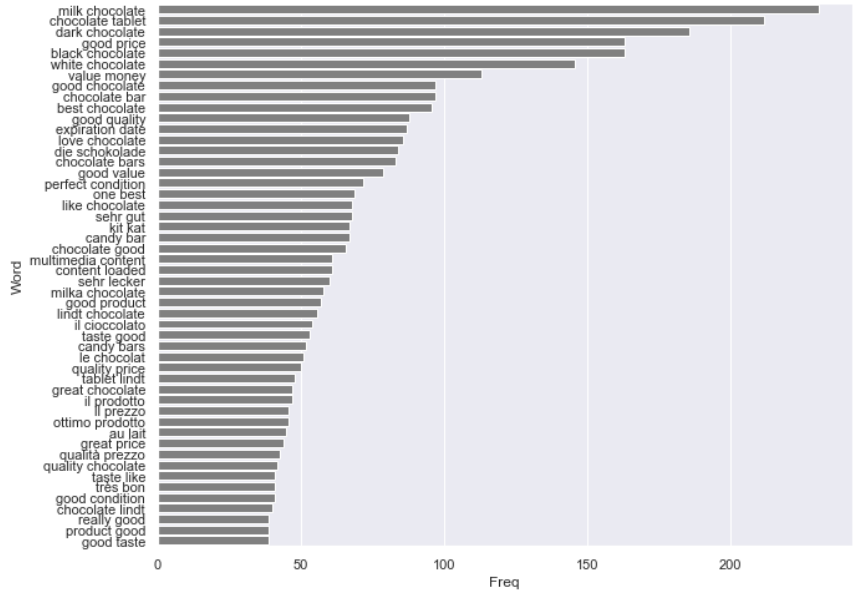

Figure 14: Bi-grams for most frequency terms.

The same happens when looking at the Tri-gram graphic in Figure 15, with terms like "good quality chocolate," "quality-price ratio," and "love dark chocolate," helping to get more specific insights based on the words that this study considers as crucial to step-up the OI Process focusing on consumers information scrapped directly from their Rating and Reviews.

```
# PLOT top 20 most frequently occuring TRI-GRAMS
top_words = get_top_n_words(processedReviews['PreProcessedText'], n=50, nGrams=3)
top_df = pd.DataFrame(top_words)
top_df.columns=["Word", "Freq"]
sns.set(rc={'figure.figsize':(10,8)})
g = sns.barplot(data=top_df, x="Freq", y="Word", color="grey")
```

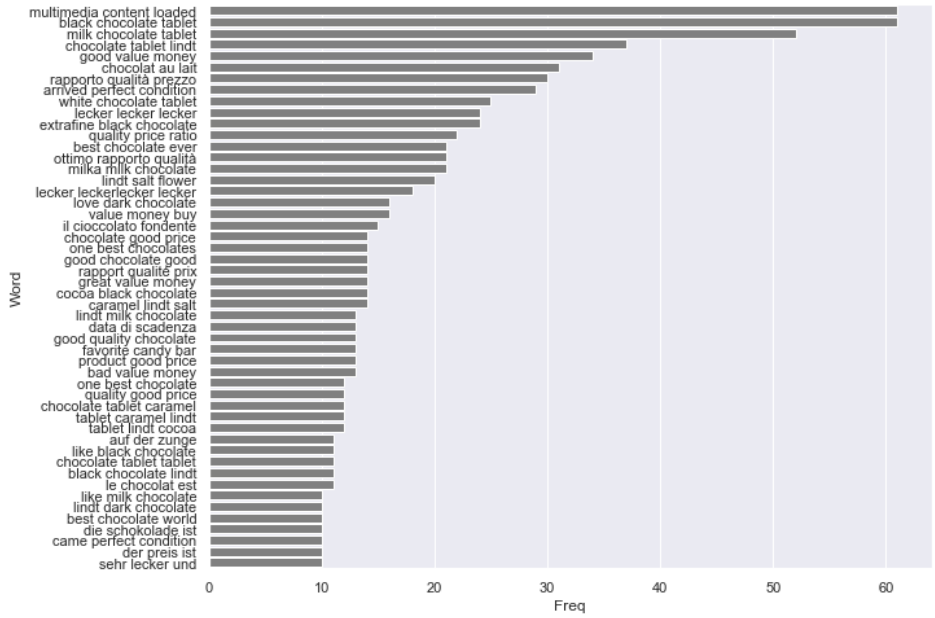

Figure 15: Tri-grams for most frequency terms.

The study could continuously analyze the composition of best terms to get more accurate insights. Still, as the purpose of this study is to understand the feasibility of gaining insights from UGC through digital asset web scrapping and if it is valuable for the OI process, it was chosen to stop the analyses in the Tri-grams.

4.4. In-Depth Research

Trying to understand how top-level executives in different innovative-driven businesses are managing new products/services within their companies, the study ran an in-depth interview aiming to collect relevant information to answer the questions: **RQ1:** What is your current methodology when driving innovation? **RQ2:** How can UGC be used to improve the OI process considering consumer-centric insights? **RQ3:** In which OI process phase should the UGC be considered? **RQ4:** How open are the senior-level executives from top Companies to using UGC in their innovative process? **RQ5:** How can UGC help to put the Consumer needs at the center of the strategy?

In this phase of the study, it was important to guarantee the avoidance of bias and have a diversification in business covered. The idea again was not to have good and usable insight but to have a flavor of how big companies are dealing with innovation instead. For that reason, the study invited 5 distinct executives from the Pharmaceutical (Pfizer EUROPE),

Coffee (Jacobs Douwe Egberts, Brazil), Research (Twomuch Research Studio, Madrid), E-commerce (Pacific Capital Partners, Portugal), and Oil and Gas industries (BR Distribuidora, Brazil) businesses, and based on a guide provided by the author, containing a high-level context and concept for all the methodologies applied, as well as first results, they were able to provide their point of view to confront the study in every topic covered.

Most of them when innovating in their companies are using the classic approach of developing new products and services. The usage of research to find macro trends, internal brainstorming and hackathons, and screening to select ideas seems to be the very first step in their innovation process. Following the traditional methodology, they also mentioned Qualitative and Quantitative research as well as a consumer panel to start validating the assumptions and see how they fit with the business objective. However, they are also concerned about the usage of the same consumer panel many times and creating a bias during the process.

This phase of the innovation process intends for the first time put the consumer at the center of the problem to be solved.

When asked about how can UGC be used in their process, the study shows the start of divergence in ideas.

All of them agreed that having a huge amount of consumer data insight with low or no cost involved is the main benefit of the methodology proposed, mainly in a very costly process, but they can also see other benefits like connecting to different sources of information anonymously, removing the bias that they usually have in their innovative process and statistically having accuracy due to the size of the sample where the most data, the better.

The divergence comes as a possible drawback of not being aware of the context when the information was collected, even using text mining and topic modeling to transform the big lake into pure raw and insightful data to be used, where classical complementary research would be needed.

Another branching found was where the UGC should be allocated in the innovation process. For some, it should be at the very beginning of the process, helping to find the initial consumer needs driving the development ahead, but for others, it can also be used across the process not only gathering insightful information but also validating the whole scheme, like content, prototypes, etc.

The common sense of this in-depth interview shows that all of the five top executives are really open to using UGC content collected through the methodology presented in this study as a way to reduce innovative costs, to detect signs of something relevant in motion, and mostly important to have reliable consumer insights when developing a product/service having them finally in the center of their strategy.

5. CONCLUSION, DISCUSSION AND FUTURE WORK.

Most companies have the consumer at the cornerstone of their strategies, but few are open to adopting methodologies that derail from the classic approach to innovation. Internal brainstorming, a panel of consumers, qualitative and quantitative research, or even in-company hackathons are used to fill consumer needs and generate insights to create a competitive edge, but consumers themselves are consulted just for validation and in a specific phase of the innovation process.

The possibility to use a large amount of consumer information available, and extract from there, useful and trustful insight can be a game changer for every company that aims to have consumer needs at the core of their strategy.

The usage of the right methodology to anonymously webscrap available data from the most relevant and reliable sources like social media, e-commerce platforms, publishers, competitors' websites, and so on, and the ability to create codes that can run good text mining as well as topic modeling including keywords extraction and frequency analyses, have the potential to generate good insights that can be used in the innovation process in an end-to-end approach.

Despite being a good solution for innovation, most companies are afraid of using OI, because it can raise a discussion about information security, ownership of the ideas generated, and governance during the process, which can be costly and time-consuming, what is the opposite of an innovative mindset, and the usage of UGC can easily cover this gap, bringing to the table good quality data from outside, without jeopardizing the process.

Important to highlight that the usage of UGC to speed up the OI adoption needs to fall under a procedure that can bring a minimum level of reliability through coherence and perplexity methodology, guaranteeing that the data, even without context can be used to make proper decisions.

Worth mentioning that new technologies like Machine Learning and Generative Artificial Intelligence, can help to create more ways to collect data and collate the information with unprecedented accuracy, which can accelerate OI adoption with a significant cost reduction.

After running a small sample of 9600 consumer Ratings and Reviews from Amazon Iberia using Amazon's SKU number (ASIN) from 2016 to 2023, the study showcases the scalability of the process, depending on the challenge ahead and the level of accuracy intended.

The study suggests for future research, ways to fine-tune the methodology using the most updated tools available at the moment, and the usage of that information not only for the private sector as there is an uncountable opportunity also for the public sector to leverage their services, reduce costs and increase their reputation, just truly listening to consumers needs and developing real consumer-based solutions.

BLIOGRAPHICAL REFERENCES

Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R and D Management*, *50*(3), 329–351. https://doi.org/10.1111/radm.12408

Artem Timoshenko, by, Hauser Kirin Professor of Marketing, J. R., Sloan, M., Tucker Sloan Distinguished Professor of Management Professor of Marketing Chair, C., & Sloan Program, M. (2017). *Identifying Customer Needs from User-Generated Content LIBRARIES ARCHIVES Signature redacted Signature redacted Signature redacted Thesis Supervisor Identifying Customer Needs from User-Generated Content*.

Audretsch, B. D., & Belitski, M. (2022). The limits to open innovation and its impact on innovation performance. *Technovation*. https://doi.org/10.1016/j.technovation.2022.102519

Barlatier, P., Josserand, E., Hohberger, J., & Mention, A. (2022). Configurations of social media-enabled strategies for open innovation, firm performance, and their barriers to adoption. *Journal of Product Innovation Management*. https://doi.org/10.1111/jpim.12647

Barrett, G., & Tsekouras, G. (2022). A tango with a gorilla: An exploration of the microfoundations of open innovation partnerships between young innovative companies and multi-national enterprises. *Technovation*, *117*. https://doi.org/10.1016/j.technovation.2022.102561

Brunetti, F., Matt, D. T., Bonfanti, A., De Longhi, A., Pedrini, G., & Orzes, G. (2020). Digital transformation challenges: strategies emerging from a multi-stakeholder approach. *TQM Journal*, *32*(4), 697–724. https://doi.org/10.1108/TQM-12-2019-0309

Cavallo, A., Burgers, H., Ghezzi, A., & van de Vrande, V. (2022). The evolving nature of open innovation governance: A study of a digital platform development in collaboration with a big science centre. *Technovation*, *116*. https://doi.org/10.1016/j.technovation.2021.102370

Colicev, A., Kumar, A., & O'Connor, P. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *International Journal of Research in Marketing*, *36*(1), 100–116. https://doi.org/10.1016/j.ijresmar.2018.09.005

Dahlander, L., Gann, D. M., & Wallin, M. W. (2021). How open is innovation? A retrospective and ideas forward. *Research Policy*, *50*(4). https://doi.org/10.1016/j.respol.2021.104218

Gupta, R., Mejia, C., & Kajikawa, Y. (2019). Business, innovation and digital ecosystems landscape survey and knowledge cross sharing. *Technological Forecasting and Social Change*, *147*, 100–109. https://doi.org/10.1016/j.techfore.2019.07.004

Ho-Dac, N. N. (2020). The value of online user generated content in product development. *Journal of Business Research*, *112*, 136–146. https://doi.org/10.1016/j.jbusres.2020.02.030

Hsiang, C. Y., & Rayz, J. T. (2022). Predicting popular contributors in innovation crowds: the case of My Starbucks Ideas. *Information Technology and People*, *35*(2), 494–509. https://doi.org/10.1108/ITP-04-2019-0171

Li, L., Zhu, W., Wei, L., & Yang, S. (2022). How can digital collaboration capability boost service innovation? Evidence from the information technology industry. *Technological Forecasting and Social Change*, *182*. https://doi.org/10.1016/j.techfore.2022.121830

Lin, J., Wang, C., Zhou, L., & Jiang, X. (2022). Converting consumer-generated content into an innovation resource: A user ideas processing framework in online user innovation communities. *Technological Forecasting and Social Change*, *174*. https://doi.org/10.1016/j.techfore.2021.121266

Lu, Q., & Chesbrough, H. (2022). Measuring open innovation practices through topic modelling: Revisiting their impact on firm financial performance. *Technovation*, *114*. https://doi.org/10.1016/j.technovation.2021.102434

Madanaguli, A., Dhir, A., Talwar, S., Clauss, T., Kraus, S., & Kaur, P. (2022). Diving into the uncertainties of open innovation: A systematic review of risks to uncover pertinent typologies and unexplored horizons. *Technovation*. https://doi.org/10.1016/j.technovation.2022.102582

Mengist, W., Soromessa, T., & Legese, G. (2020). Ecosystem services research in mountainous regions: A systematic literature review on current knowledge and research gaps. Em *Science of the Total Environment* (Vol. 702). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2019.134581

Ogink, R. H. A. J., Goossen, M. C., Romme, A. G. L., & Akkermans, H. (2022). Mechanisms in open innovation: A review and synthesis of the literature. Em *Technovation*. Elsevier Ltd. https://doi.org/10.1016/j.technovation.2022.102621

Osorno, R., & Medrano, N. (2022). Open Innovation Platforms: A Conceptual Design Framework. *IEEE Transactions on Engineering Management*, *69*(2), 438–450. https://doi.org/10.1109/TEM.2020.2973227

Puliga, G., Urbinati, A., Franchin, E. M., & Castegnaro, S. (2022). Investigating the drivers of failure of research-industry collaborations in open innovation contexts. *Technovation*. https://doi.org/10.1016/j.technovation.2022.102543

Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2022a). Exploring the boundaries of open innovation: Evidence from social media mining. *Technovation*. https://doi.org/10.1016/j.technovation.2021.102447

Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2022b). Exploring the boundaries of open innovation: Evidence from social media mining. *Technovation*. https://doi.org/10.1016/j.technovation.2021.102447

Saura, J. R., Reyes-Menéndez, A., Dematos, N., Correia, M. B., & Álvarez-Miranda, E. (2021). Identifying Startups Business Opportunities from UGC on Twitter Chatting: An Exploratory Analysis. *J. Theor. Appl. Electron. Commer. Res*, *16*, 1929–1944. https://doi.org/10.3390/jtaer

Yildirim, E., AR, I. M., Dabić, M., Baki, B., & Peker, I. (2022). A multi-stage decision making model for determining a suitable innovation structure using an open innovation approach. *Journal of Business Research*, *147*, 379–391. https://doi.org/10.1016/j.jbusres.2022.03.063

Zhang, H., Lin, Q., Qi, C., & Liang, X. (2022). The effects of online reviews on the popularity of user-generated design ideas within the Lego community. *European Journal of Marketing*. https://doi.org/10.1108/EJM-10-2021-0816

Zhang, X., Chu, Z., Ren, L., & Xing, J. (2023). Open innovation and sustainable competitive advantage: The role of organizational learning. *Technological Forecasting and Social Change*, *186*. https://doi.org/10.1016/j.techfore.2022.122114

REMOTE INTERVIEW VIDEOS AND TRANSCRIPTIONS

Interview Guide: InterviewsGuide.pdf

Pharmaceutical (Pfizer EUROPE): Pfizer

Coffee (Jacobs Douwe Egberts, Brazil): Jacobs Douwe Egberts

Research (Twomuch Research Studio, Madrid): Twomuch Research Studio

E-commerce (Pacific Capital Partners, Portugal): Pacific Capital Partners

Oil and Gas industries (BR Distribuidora, Brazil): BR distribuidora