UNIVERSITAT DE BARCELONA

Digital Content Marketing Strategies for resource optimisation in e-Government platforms

Anabel Guzmán Ordóñez

PhD in Business

Digital Content Marketing Strategies for resource optimisation in e-Government platforms

Anabel Guzmán Ordóñez

2023

PhD in Business | Anabel Guzmán Ordóñez

UNIVERSITAT DE BARCELONA

PhD in Business

Thesis title:

Digital Content Marketing Strategies for resource optimisation in e-Government platforms

PhD candidate:
Anabel Guzmán Ordóñez

Advisor:
Francisco Javier Arroyo Cañada

Date:
July 2023

UNIVERSITAT DE BARCELONA

Dedicated to those who never clipped my wings

but propelled me to fly higher.

Acknowledgement

Expressing my gratitude to all who have supported me along this arduous research path is a privilege. First, I express my deepest gratitude to my supervisor, Dr Xavier Arroyo, for his expert guidance, patience, and dedication throughout the process. I would also like to thank my research colleagues and the professors in marketing for their valuable support, exchange of ideas and enriching debates. Your constructive feedback has contributed significantly to developing and improving this doctoral thesis.

I cannot overlook thanking my family and my love, George, for their unconditional emotional support and constant encouragement. Your love and understanding in the most challenging times have given me the strength to overcome obstacles and persevere in achieving my goals.

Finally, I thank all the institutions and companies who generously collaborated by providing the information and resources necessary to carry out this study.

My sincere thanks!

Anabel Guzmán Ordoñez

Content

List of Figures

List of Tables

Chapter 1

Introduction

Introduction

1.1 Research motivation and theoretical background

Due to advances in information and communications technologies (ICTs) made over the past two decades, governments have used them in an accelerated and substantial way to accelerate, optimise, flex, streamline and make transparent public sector processes or activities (Adjei-Bamfo, Maloreh-Nyamekye and Ahenkan, 2019). In this way, governments today are making greater use of digital channels, which they have been prioritising with communication strategies that generate more significant interaction with citizens (Torres et al., 2021). Websites and social networks, such as Facebook and Twitter, the leading digital media (Dwivedi et al., 2017), disseminate information content, statistical data, and visibility to government projects, online transactions or open data. As of January 2022, internet users worldwide were 5.16 billion, about 64.4% of the world's population (Datareportal, 2023). In Latin America, the number of internet users was 453 million, about 66.6% of the population (Datareportal, 2022). It is important to note that these numbers are constantly changing as more people gain access to the internet.

In addition, public bodies have had to adapt to emerging changes and generate strategies. One of the challenges they have faced is the Electronic Government (EG), which bases its application on public administration. It aims to contribute to using ICTs to improve the services and information offered to citizens and organisations, improve and simplify the processes of institutional support, and facilitate the creation of channels that increase transparency and citizen participation (Naser and Concha, 2011). Therefore, during the COVID-19 crisis, for instance, governments have been using information and communication technologies (ICTs) with different approaches and purposes (Criado, Guevara-Gómez and Villodre, 2020). Finquelievich and Prince (2013) claim that e-Government programs not only cover the use of ICTs but also seek to build and maintain networks between different levels, i.e., between rulers and the governed and different social actors interacting with each other. This relationship must narrow the gap between society and the state entity, generating new dynamics of participation, where citizens have more open and direct channels of communication towards public entities.

Different state public policies have been developed around the world that allow citizens to participate and have access to up-to-date information, as well as shorten the time spent on civil proceedings and online tax payments, and improve conditions for state employees in user care and paperwork, who must understand the changes in work processes that have been redesigned (Contreras Orozco, 2017). For instance, the United Nations has developed a global indicator, the United Nations e-Government development index (EGDI), which measures the countries' use of information and communications technologies to provide public services. This index is measured by three variables: The Online Services Index, the TII-Telecommunications Infrastructure Index, and the HCI-Human Capital Index (United Nations, 2022).

EG policies have been more rigorous for Latin American countries in the past ten years. New technologies have played a vitally important role in government efforts to coordinate the response to the pandemic and ensure public collaboration during this unprecedented crisis (United Nations, 2022) The United Nations E-Government Survey 2022 indicates a positive global trend towards higher levels of e-government development as countries in all regions are increasingly embracing innovation. The Survey found that citizens benefit from more advanced e-service delivery, better access to information, and other e-government services. In Latin America and the Caribbean, 58% (19) of the 33 countries surveyed exceed the global EGDI average, with six countries ranked in the "Very High" category Mexico, Grenada, Bahamas, and Colombia are in the highest rating class of the high EGDI group and are well positioned for e-government development (Caribbean, 2022). The proportion of countries in the high levels has been increasing progressively since the first editions of the Survey. Líppez-De Castro and García Alonso (2016) affirmed in their study of citizen participation in websites that ICT implementation policies in Colombia have been advanced, such as websites that are used as channels to encourage participation in the decision-making process, and the provision of tools, such as information provision, online consultations, active participation or citizen engagement.

This use of new technologies, especially social media, has played a vital role in government efforts to coordinate responses to extraordinary events such as the pandemic and ensure public collaboration during this unprecedented crisis (Criado, Guevara-Gómez and Villodre, 2020). These social networks have played a critical role in maintaining the functioning of societies during prolonged periods of lockdown, as well as supporting solutions across sectors and national borders (Chen et al., 2020). Government entities' use of these social networks presents a unique opportunity to approach citizens but also raises challenges and issues related to the transparency of information, the measurement of the use of these platforms and engagement rates (Santoso, Rinjany and Bafadhal, 2020).

Digital content marketing (DCM) plays a fundamental role in the context of public entities or governments, providing several benefits and opportunities in digital channels like social media. First, it allows these entities to establish effective communication with citizens, disseminating relevant and up-to-date information on government policies, programs, services and events (Henisa and Wilantika, 2022). This DCM approach provides a direct channel for message delivery and interaction with the target audience. In addition, DCM in social networks plays a prominent role in promoting citizen participation (Choi and Song, 2020). Government entities can use content strategies to foster citizen collaboration and engagement, inviting citizens to participate in surveys, public forums, and other consultation mechanisms (Santoso, Rinjany and Bafadhal, 2020). This empowers citizens and allows government entities to make more informed decisions representing community needs and wants.

Transparency and accountability are essential elements in effective governance, and social media provides an appropriate platform to foster these principles (De Blasio and Selva, 2018). Government entities can use content strategies to publish reports, statistical data, budgets, and management results, promoting openness and trust in public institutions. The availability of government information on social media allows citizens to access relevant data directly, contributing to greater transparency and strengthening the link between government and society (Rumbul, 2016). Furthermore, DCM on social networks facilitates the education and awareness of users on issues relevant to the community (Hollebeek and Macky, 2019). Government entities can use content strategies to disseminate informational and educational information, such as videos, infographics, and articles, addressing health, safety, environment, and civil rights issues. This dissemination of content allows to increase the community's knowledge and combat the scepticism related to advertising and other forms of traditional communication (Holliman and Rowley, 2014).

On the other hand, using these networks with different intentions generates a need to measure effectiveness by different means and methodologies like the engagement rate (Bonsón, Perea and Bednárová, 2019). Digital channels, by nature, are a source of much information that should be used better. This paper seeks to measure the effectiveness of the content generated by public entities in social networks such as Twitter, Facebook and Instagram to identify which factors of most significant incidence have to improve the DCM strategies used.

1.1.1 Research Questions

RQ1: What is the state-of-the-art digital content marketing strategies that are used on e-Government platforms?

RQ2: What are the digital content marketing strategies used in Twitter that help with resource optimisation?

RQ3: What are the digital content marketing strategies used on Facebook and Instagram that help with resource optimisation?

1.1.2 Ethical Issues

The present study has used web scraping techniques to extract data from social networks, specifically Twitter, Facebook and Instagram. In carrying out this process, various ethical considerations were taken into account to ensure respect for the fundamental principles of privacy and Data processing. Firstly, the terms and conditions of the social media platforms, ensuring compliance with all established restrictions and policies, were reviewed (META, 2022; TWITTER, 2022). Work has been carried out within the limits permitted by these platforms, and the regulations applicable to online data collection have been respected. In addition, special attention has been paid to the privacy of users. No information or personal data has been collected from people or users of social networks other than the content extracted from profiles of public entities.

Finally, it is essential to note that the use of data obtained through web scraping has been strictly limited to this study's academic and research purposes. Results and conclusions are presented in aggregate form, and no specific individual is identified.

1.2 Objectives

1.2.1 General Objective

To define the digital content marketing strategies for resource optimisation on e-Government platforms.

1.2.2 Specific objectives

- To identify the digital content marketing strategies (DCMS) used on e-Government platforms and create a bibliometric analysis and literature review.

- To analyse the digital content marketing strategies (DCMS) used on Twitter for government entities.
- To analyse the digital content marketing strategies (DCMS) used on Facebook and Instagram for government entities.

1.3 Methodological Approach

Content creation for citizen participation and information generation occurs in a social environment and depends directly on the people who can create and develop each strategy (Bonsón, Perea and Bednárová, 2019). The measurement of the interactivity of users in social networks has been decisive in measuring the effectiveness of content (Muñoz-Expósito, Oviedo-García and Castellanos-Verdugo, 2017). Many of these digital channels have different ways of presenting content, so measuring their interactivity with users varies. Over the last ten years, many studies have presented different ways of measuring this interactivity, from traditional extractions of social network metrics to the present day, where automated methods have allowed to obtain a greater scope in the investigations (Bonsón and Ratkai, 2013; Haro-de-Rosario, Sáez-Martín and del Carmen Caba-Pérez, 2016; Bonsón, Perea and Bednárová, 2019; Henisa and Wilantika, 2021).

In this way, social networks as sources of data extraction offer a wide variety of content, metrics and high-quality information to investigate interactivity, perceptions and user behaviours. For this, the methodological approach of this thesis focused on the use of technologies and data science as the primary source for data collection, processing and analysis. In the digital marketing environment, a fundamental challenge is evaluating a strategy's success (Hollebeek and Macky, 2019). Therefore, key performance metrics used to measure methodological models and approaches should be selected and understood (Saura, 2020). The use of data science in this study allowed a deeper understanding of phenomena and variables that directly affect the effectiveness of the content, as well as measurements that were not previously considered. In addition, using data science helped improve the accuracy and objectivity of research results. By using data-driven approaches, it was possible to reduce subjectivity and bias and strengthen the validity and reliability of the conclusions (Shah and Murthi, 2020).

The methodologies used for the development of this thesis were segmented for each of the objectives. First, a bibliometric study was carried out to achieve a greater scope in verifying the state of the art of literature. This bibliometric study as a research methodology allowed to analyse and evaluate quantitatively the scientific production in the field of study by collecting bibliographic metrics and citations in databases (Donthu et al., 2021) such as Scopus and Web of Science. Likewise, the execution of this methodology provided a global perspective of state of the art, in addition to identifying the areas of knowledge that require more attention, generating innovative ideas for future research and proposing contributions to the area in research better contextualised based on the most significant findings (Merigó and Yang, 2017).

On the other hand, it was necessary to include agile methodologies in data analysis for the analytical model proposed in this thesis. After evaluating different characteristics, the CRISP-DM (Cross-industry Standard Process for Data Mining) was taken as a reference framework for this type of research (Palacios et al., 2016). This methodology is divided into six phases: understanding the business, understanding the data, modelling, evaluating and deploying (Wowczko, 2015). Each phase included data processing and analysis activities to obtain results that gave deeper scope in the study area.

Finally, developing research based on agile methodologies and data allowed the fulfilment of the proposed objectives to be achieved optimally and with a greater scope than thought at the beginning of the study.

1.4 Thesis Structure

The thesis was structured to respond to the main objective, specific objectives, and previously asked research questions. For this purpose, this response will be developed in chapters two (2), three (3) and four (4), as specified in the following Table 1. with each of the topics considered. It should be noted that chapters one (1) and five (5) (Introduction and Conclusions) are typical of the general process of the research and do not go into detail as such in the main structure.

Table 1. Thesis Structure

Chapter	Title	Methodology	Main findings
Chapter II.	Literature review of content marketing strategies for e-Government platforms	Bibliometric analysis	State of Art
Chapter III:	Data collection in content marketing strategies for e-Government platforms	Databases created from extracting content from social networks (Twitter, Facebook and Instagram).	Analytical model for ERP in Twitter
Chapter IV:	Structuring a model of content marketing strategies for e-Government platforms	Machine learning model to generate the best results for content marketing strategy	Analytical model for ERP in Facebook and Instagram

Chapter Two (2) develops the first study of the literature review. The primary purpose is to investigate the current evolution and state of the literature by using a quantitative-descriptive methodology and to analyse the metadata that the database yields. Therefore, the study uses a bibliometric analysis which initially examines knowledge development in a specific area or topic, scientific quality and influence of works and sources quantitatively (Merigó et al., 2015). To evaluate the quantitative criteria, the search selected indicators specific to the bibliometric analysis (Donthu et al., 2021). Among these are the number of scientific productions in the study area, the number of publications by authors, and the number of citations per article. Likewise, the Web of Science (WOS) and Scopus platform yields results by scientific category (Business, Management, Communication, Computer Science and Information Systems), the evolution of publications by year, the types of published documents (articles, book chapters, conference proceeding papers, early access or reviews), as well as educational organisations or institutions, authors, countries or regions, publishers, languages, research areas, among others. This facilitates the analysis of scientific production in this field and contributes to the development of the first specific objective of this study.

Chapter Three (3) aims to develop an analytical model that identifies the content characteristics that enhance the effectiveness of Twitter accounts. The study uses the governorates of Colombia as a case study. It employs the CRISP-DM methodology for data mining to clean, process, and analyse all data collected from the accounts of Colombian governments. The results indicate that variables such as the number of followers and the publication cadence in the accounts do not improve Engagement Rate per Post (ERP) rates, but the content type does. The model identifies the variables that improve the ERP and their impact on the effectiveness of the content.

Finally, Chapter Four (4) examined the effectiveness of content marketing on Facebook and Instagram, focusing on the context of Colombia's governorates through an analytical approach. The proposed methodology involves collecting and analysing data related to the content published by the governorates on these platforms, as well as the interaction and participation of citizens using CRISP-DM. Key indicators were considered, such as the reach of publications, user interaction and the impact of content marketing campaigns on promoting services, programs, and government policies through shared content.

Chapter 2

A Bibliometric Review

A Bibliometric Review of Digital Content Marketing Strategies for Electronic Government: Trends and Insights

1. Introduction

Nowadays, Digital channels are saturated with content shared chiefly organically by users of social networks, websites, or mobile applications, whether to search for product or service information, purchase references from other users or consume written visual or auditory content. The customer-to-customer interaction in digital environments has affected users' attitudes, preferences, and consumption behaviour (Berger and Milkman, 2012) for decades. Internet-based technology has caused marketing to encourage consumer participation (Vinerean, 2017). Driven by a transformation of brands and organisations due to content generation, nine out of ten companies use content marketing to capture, interact, convert, and remind consumers about products and brands (Jutkowitz, 2014). Therefore, content marketing changes how companies sell and communicate with their target audience (Kee and Yazdanifard, 2015). Consumers are no longer satisfied with just knowing the existence of a product or service. They already find more detailed information before purchasing and looking for brands that generate value for them (Hutchins and Rodriguez, 2018; Schultz, 2016). This forces organisations to have content strategies relevant to the target audience, attract new customers non-intrusively, and increase their interaction (Deighton and Kornfeld, 2009).

In 2020, the development of the pandemic due to COVID-19 accelerated the adoption of digital technologies in areas where development had been delayed or governments and companies had not prioritised, such as data collection techniques, telemedicine, home office, e-learning, and use of digital platforms (Hantrais et al., 2020). Therefore, and taking advantage of this trend, governments sought to use this data and digital technologies to go further: to promote new forms of government that are more participatory, innovative, and agile (Criado, Guevara-Gómez and Villodre, 2020). Digital governance involves the complete digitisation of the public sector, enabling the level of integration needed to deliver better services to citizens and businesses (Welby, 2019). These practices facilitate transforming services and collaboration in public sector organisations, making them more open, user-oriented, and proactive. Well-established digital governments can help make the government more resilient and responsive (Welby, 2019), which has become extremely important in times of emergency, as the COVID-19 pandemic has shown.

Currently, Internet users worldwide are 4,660 million people, 316 million more than in January 2020, representing 7.3% to January 2021 (Kemp, 2021). Likewise, the use of social networks has also impacted the number of registered users. According to figures from the Digital Global Overview Report, 4,200 million users worldwide, 490 million more than the last 12 months, representing year-on-year growth of 13%" (Kemp, 2021). Despite the changes generated by COVID-19 in the behaviours of users concerning the consumption of digital content, the same study showed that the time spent on social networks is approximately the same as in the year 2019, which is 2 hours and 25 minutes each day, demonstrating that it has only increased on average 30 minutes for the data of the last five years. The government's use of these digital media has caused users to believe in the influence of improving or worsening the transparency of information transmitted through social networks (Sweeney, 2019).

Hence, open government can be conceived as a policy-oriented to transparency, participation and collaboration supported by digital technologies to achieve its objectives (de Blasio and Selva, 2019). Transparency involves horizontal accountability mechanisms (Mulgan, 2014), such as disclosing open data, providing tools to discuss and debate with administrators, and extending citizens' powers of control and enforcement. Participation can be achieved to increasing degrees (de Blasio and Sorice, 2016); public consultations are only the first step to a fully collaborative decision-making process (de Blasio and Selva, 2019). According to the study presented by CIGI and IPSOS (2019), countries such as Kenya, Indonesia and Nigeria have an average percentage of 60% where users consider that government transparency has been improved thanks to social networks. On the other hand, in countries such as Japan, the United Kingdom and the United States, 64% of users consider that there is no more significant impact on the transparency of social networks.

The scientific activity of recent years allowed this study based on bibliometrics to be carried out since the considerable increase in research in business, marketing, and the electronic government has been reflected in the results of searches made in the Databases of Web of Science and Scopus. Thus, the objective of this study was to investigate the evolution and current state of the literature on both terms related to content marketing and digital governance based on a bibliometric analysis, which allows an assessment of the scientific activity and the impact of the research and sources. In addition, bibliometric analysis, compared to the systematic literature review, allows us to see more broadly the state of the art of the topics investigated and reflect emerging trends in a subject or field of research (Donthu et al., 2021). Furthermore, using a quantitative-descriptive methodology and analysing the metadata was possible to determine the relationships between terms and keywords to discern advanced research on these topics.

The main results obtained are in the use of keywords such as "Social Media", "e-Government", and "Digital Government" within most of the literature consulted, in addition to others related such as "Political Communication", "Media", "Engagement" and recurrent use of words such as "Twitter" or "Facebook" as primary communication channels for government entities.

1.1 Digital Content Marketing Approach

Academic interest in researching Digital Content Marketing (DCM) topics is also increasing rapidly. Since 2008, Rowley (2008) has stated that DCM is the responsible management process for identifying, anticipating, and satisfying the requirements of digital users and those who are distributed through digital means. Pulizzi (2012), on the other hand, defines DCM as a relevant and compelling creation that adds value and is used to produce positive user behaviour or a brand perspective. In addition, DCM commercialisation includes creating, organising, distributing, and expanding content that is interesting, relevant, and useful to clearly defined audiences to create discussions about content that has already been shared and distributed (Kotler, Kartajaya and Setiawan, 2017). Handley and Chapman (2011) define content as anything created and uploaded to a website: words, images or other existing things. Halvorson, Rach and Cancilla (2012) suggest that content is what the user came (to your website) to read, learn, watch or experience.

Wuebben (2011) sees content as the key component to telling a brand's story, the story of your product or service and drives your brand into the hearts and minds of your potential consumer, clients and leads. Other authors (Halvorson, Rach and Cancilla, 2012; Scott, 2020; Bloomstein, 2012) suggest variations that give a slightly different approach. Silverman (2012) concludes that the purpose of content marketing is to attract potential customers and complement the brand's credibility, while Godin (2007) comments that content marketing "is the only marketing that remains. However, from a relationship marketing perspective, a pertinent question is how this consumer engagement with content (as opposed to the brand) evolves into brand engagement, moreover, how this brand engagement can be fostered adequately through the same digital content marketing interactions (Opreana and Vinerean, 2015) to generate changes in positive attitudes (e.g., trust in the brand) or behaviours (e.g., brand-related interactions) in the users (Koob, 2021).

DCM as a digital marketing strategy is part of methodologies that have emerged in recent years, as it is Digital Inbound Marketing (DIM). Opreana and Vinerean (2015) define the DIM as reaching and converting qualified consumers by creating and pursuing organic tactics in digital environments. Patrutiu-Baltes (2016) states that the DIM is a strategy for connecting with potential customers through functional materials and experiences to them. Using media such as blogs and social networks, marketers expect to entertain and inform viewers of the content they seek for themselves. This methodology is based on adding value in the four phases of the entire customer journey funnel: attract, convert, close and delight (Samsing, 2018).

In this way, the DIM does not need to fight for the attention of potential customers. By creating content that addresses the problems and needs of ideal customers, the brand attracts qualified leads and builds trust and credibility (Dakouan and Benabdelouahed, 2019). The content becomes a key component of DIM; its use in organisations attracts potential customers, retains existing ones, and makes consumers aspirational by developing numerous content that potential customers value from the brand (Opreana and Vinerean, 2015).

These approaches are widely applicable, both in business-to-business (B2B) models (Taiminen and Ranaweera, 2019; Galvez Torres et al., 2020) and in organisations that have a direct relationship with the customer, business-to-clients (B2C) (Zhang and Du, 2020). Furthermore, the DCM strategy can be extended to diverse industry sectors, like private and public, especially in digital governance strategies. The use of digital platforms by government bodies for digital content distribution has increased, informing and allowing citizen participation more directly and transparently (Sweeney, 2019).

Therefore, governments have accelerated ICT substantially to streamline, optimise, and make flexible and transparent public sector processes or activities (Naser and Concha, 2011). In this way, public bodies use digital channels more, guided by an accelerated digital transformation. They have prioritised communication strategies that generate more significant interaction with citizens (Sweeney, 2019). Websites and social networks like Facebook and Twitter are the leading digital media (Dwivedi et al., 2017). They are used to disseminate information content, statistical data, visibility of government projects, online transactions, and open data and encourage users to communicate directly (Dwivedi et al., 2017; Sweeney, 2019).

2. Methodology

For the present study development, a systematic search was carried out to determine the characteristics of the development of literature and data analysis based on a bibliometric search since this provides a general image of the research field through bibliographic material and allows classification by categories such as articles, authors, and journals (Merigó and Yang, 2017). The methodology is based on mixed research methods, which seek an approach to the reality, the relationships, strategies, and techniques that will be used and established through a preliminary design (Ramírez-Montoya and Lugo-Ocando, 2020). These methods are usually defined as combining different research methods or multiple methodological strategies to study and answer questions about a particular topic. Among the definitions of mixed methods, Plano-Clark and Ivankova (2016) conceptualise them as the intentional integration of quantitative and qualitative research approaches to address a research problem better. Meanwhile, others have defined them as the ability to perform balanced study analyses that increase the validity of a justification (Edmonds and Kennedy, 2017) and its scope (Onwuegbuzie and Teddlie, 2003).

Initially, the bibliometric analysis examined the quantitative knowledge development on these specific topics, the scientific quality, and the influence of works and sources (Merigó, Mas-Tur, Roig-Tierno and Ribeiro-Soriano, 2015). For the evaluation of the quantitative criteria, in the search, indicators specific to bibliometric analysis were selected (Sancho, 1990; Fernández and Bueno, 1998). Among these are the number of scientific productions, the number of publications per author, and the number of citations per article. Likewise, the Web of Science (WOS) platform yielded results by scientific category (Business, Management, Communication, Computer Science, Information Systems), the evolution of publications by year, types of documents published (articles, book chapters, conference proceeding papers, early access or reviews), organisations or educational institutions, authors, countries or regions, publishers, languages, and research areas.

The criteria used in this study for the evaluation of the selected sample are directed by the bibliometric indicators found in the content marketing research field, social media and digital marketing given by authors such as Van Osch and Coursaris (2014), which highlight their importance in the state and future trend of the research area. In addition, previous scientific studies were taken as a reference (Dagnino, Levanti, Minà and Picone, 2015; Kumar, Sharma and Salo, 2019; Paesbrugghe, Sharma, Rangarajan and Syam, 2018; Randhawa, Wilden and Hohberge, 2016) which suggest within the methodology of bibliometric analysis four stages of analysis:

1. Sample selection and citation analysis.
2. Co-citation analysis to identify subareas in the research.
3. The text analysis is to understand the change in the literature.
4. The text analysis identifies trends that help in the future research direction.

This complement to quantitative analysis is given to understand the study better since few studies combine the two methods, both qualitative and quantitative, in topics related to content marketing. However, it is necessary to demonstrate the evolution of the literature and determine limitations or gaps to propose a way forward (Bhatt, Ghuman and Dhir, 2020).

2.1 Keyword selection and sample definition

The Web of Science (WOS) database was searched to compare the results, determine the amount of academic output over the past ten years, and assess the evolution of the literature. The formula used considers keywords related to "content marketing" and "digital government" or "e-government".

TEMA: (*egovernment* or *e-government* or *digital government*)

AND TEMA: ("*social media*" or "s*ocial marketing*")

AND TEMA: (*content marketing* or *content strategy*)

Once the keywords and methodology have been chosen, the initial systematic search includes the most relevant scientific papers. Due to WOS's significance and the bibliometric indicators generated, it was chosen as the database consulted for this bibliometric study. The platform provides a detailed search analysis and collects many indexed journals of any discipline and especially the object of study such as Marketing, Management, Communication and Business, as well as open access to indexed content such as the Directory of Open Access Journal (DOAJ) and the Scientific Electronic Library Online (SciELO). The search conducted in WOS displayed 210 articles. As a result, the search was examined, and 147 articles met the research theme.

On the other hand, WOS calculates the impact factor that scientific publications have, quantifying the number of times an article is cited in each period and the number of articles published by a journal. This way, scientific publications are positioned, and their relevance within the research field is evaluated. This classification is determined by the Institute for Scientific Information (ISI) in its annual report known as the Journal Citation Report (JCR), where the impacts of scientific publications are collected according to criteria of quality and scope. In Marketing and Communication, the Social Science Citation Index (SSCI) brings together the best publications in the social sciences (Bhatt et al., 2020).

2.2 Data selection and extraction process

The documents that were taken for the study were articles from indexed journals which, in their title, abstract, or list of keywords, were related to "content marketing", "social media", "e-government", and "digital government", in a search range of ten years (2011-2020), with open access and considering two languages both English and Spanish.

Articles that were not related to search keywords or had at least one but did not connect to the purpose of the study were not considered. Likewise, chapters of books and articles written in languages other than English or Spanish and not in the time range (2011-2020) were excluded.

2.3 The data organisation in preparation for analysis

The sample of 147 articles was processed considering each bibliometric category that WOS throws, like the total number of publications per year, number of publications by research area, geographical area or country, most used languages, institutions, and journals with more publications related to these topics. In addition, the content was categorised with a frequency map, seeking to identify the connectivity between these terms. Through clusters of words, it was possible to identify the potential gaps existing to consider them in future research. Studies such as those carried out by Valenzuela et al. (2017) and Martínez, Merigó, Valenzuela and Nicolás (2018) were taken as a reference in the methodology of data collection and analysis of the leading bibliometric indicators.

3. Results and Analysis

The research results were analysed in two phases. In the first phase, the qualitative analysis was done, where the 147 articles were taken and processed in the VOSviewer tool to determine the relationship between the terms found. In the second phase, the bibliometric data of Web of Science (WOS) related to the search topics were exported and analysed to determine characteristics of the scientific activity related to topics of "digital content marketing" and "digital government" in conjunction with the complementary data for each category.

3.1 Analysis of the co-occurrence of scientific documentation

As an initial result, 692 keywords were identified in the 147 articles analysed, of which 32 were prioritised. The map was elaborated (Figure 1) with the co-occurrence of terms and the multivariate clustering analysis statistical technique (Gálvez, 2018).

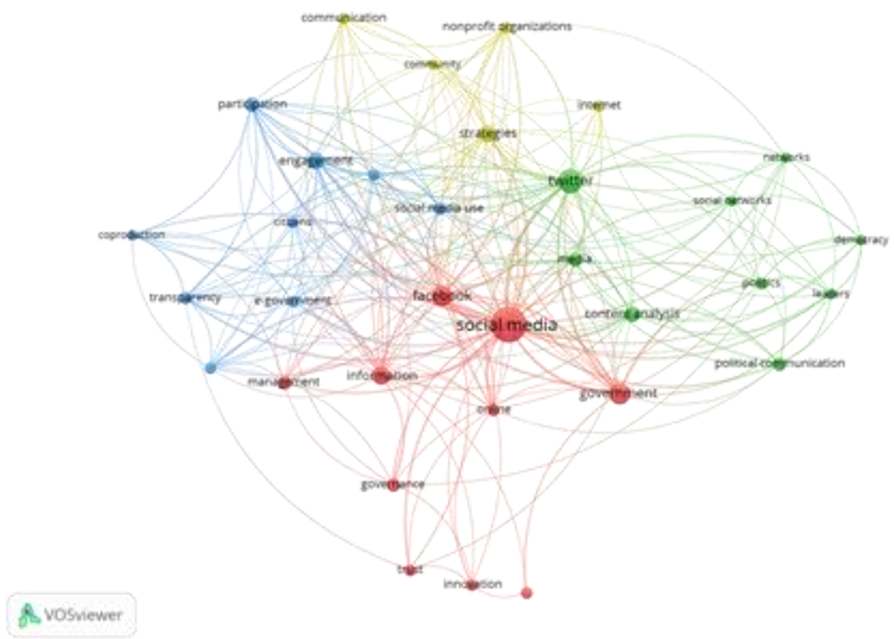


Figure 1. Terms of co-occurrence map

The terms with the highest co-occurrence were "Social Media", with a relative frequency of 36, followed by "Twitter", with 21, "Government", with 17 and "Content Analysis", with 9. Although the last term is not followed in values of the previous one, it is relevant for this study to know the relationship between the previous three. The term "Social Media" (Figure 2a) presents a relationship with terms such as "Facebook", "Twitter", "Government", and "Participation" being the one that has the most incidence and frequency. On the other hand, the term "Government" (Figure 2b) is related to terms such as "Social Media", marking that incidence that was shown in Figure 2a, followed by terms such as "political communication", "transparency", "Twitter" and "Facebook", evidencing that these social networks are the most studied by the academy to relate digital government issues, communication and participation channels.

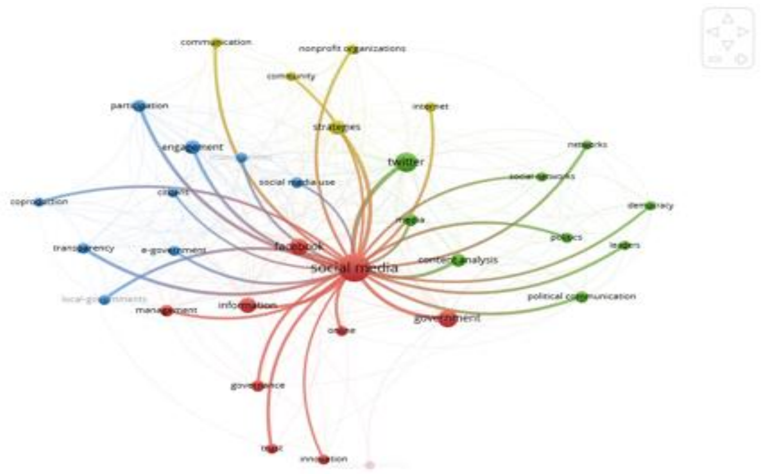
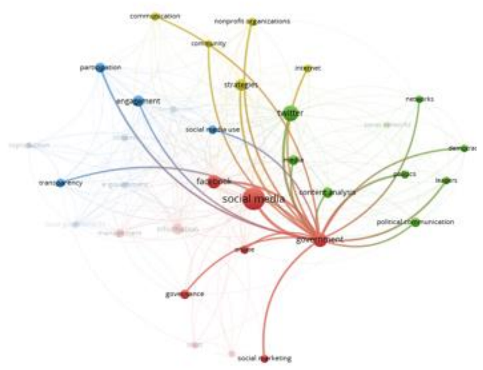


Figure 2a. “Social Media” cluster



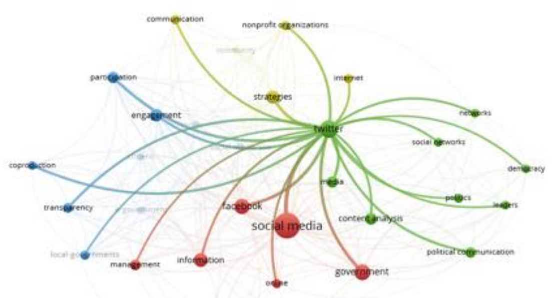


Figure 2b. “Government” cluster

Figure 2c. “Twitter” cluster

Figure 2. Keyword cluster map

The keywords cluster related to "Twitter" (Figure 2c) shows a more significant co-occurrence with the word "government" concerning the other terms, being the most used social network to talk about political issues and issues such as transparency in information. The cluster formed by the term "content analysis" was related to Facebook, Twitter, and Political Communication. There was no co-occurrence with the term "strategy" or "content marketing", evidencing a gap in the literature and thus showing an opportunity for future studies.

3.2 Analysis of bibliometric data

In the last ten years, there has been an increase in scientific production and interest in researching digital content marketing and digital government topics. For 2018 and 2020 (Figure 3), the increase was three times greater than in 2017, with an annual average of 37 and 11 articles registered, respectively.

These data are related to technological advances and the accelerated adoption of digital communication channels such as websites and social networks by government bodies and public entities (Naser and Concha, 2011; Dwivedi et al., 2017; Sweeney, 2019). For the object of the study, two searches in the WOS database were done, at the beginning of 2020 and the second in October, to determine if the pandemic had any impact on scientific production. Moreover, the increase in academic production is shown in Figure 3. The relationship between DCM and digital government themes with COVID-19 was constant in papers related to these topics, besides how the pandemic accelerated the digital transformation in public bodies.

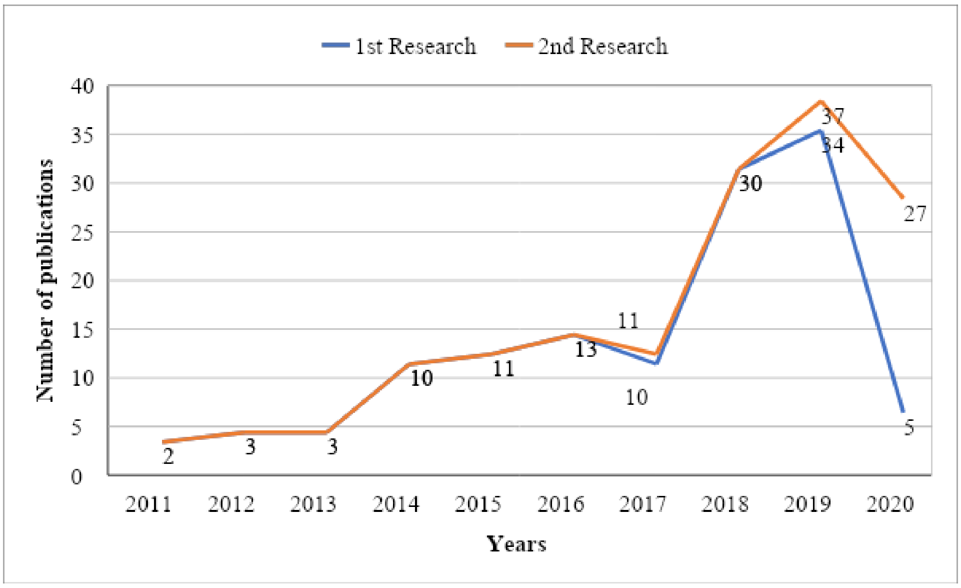


Figure 3. Publications per year on the Web of Science -WOS

On the other hand, the data also showed the main research areas that have contributed to DCM and digital government (Table 1). These were Information Science with a significant percentage of 15.6%, followed by Business Economics with 12.9% of publications concerning the sample, the 147 articles being analysed. Third place is Government Law with 10.9% and Social Science with 10.7%. Research areas such as Computer Science, Public Environmental, and Occupational Health represent 8.2%, including several publications of 12 articles annually. Finally, Public Administration, Health Care Sciences Services, and Sociology have a slightly lower percentage of participation, but that shows the interest of researchers from other areas to produce information on topics such as content marketing, social media, and digital government.

Table 2. Number of publications by research area (2011-2020)

Research Area	NP	PC
Information Science Library Science	23	15.6%
Business Economics	19	12.9%
Government Law	16	10.9%
Social Sciences Other Topics	13	10.7%
Computer Science	12	8.2%
Public Environmental Occupational Health	12	8.2%
Public Administration	9	6.1%
Health Care Sciences Services	8	5.4%
Sociology	6	4.1%

Notes: NP= Number of publications; PC= Percentage of contribution

Therefore, Table 2 shows the twelve primary sources with the highest contribution to DMC and digital government, the bibliometric data to be considered where the number of publications (NP), the significant percentage (PC), the country of origin (CO), the impact factor of the source, measured by the Journal Citation Report (JCR) and the SCImago Journal Rank (SJR). Initially, it can be evidenced in each source that there is dispersed participation in the scientific production of topics such as content marketing for digital governments. Journals can be categorised into three main themes: government and public administration, technology and areas of information, business, and marketing. Most sources come from countries such as the United Kingdom, the United States, Canada, and Spain.

The source with the most outstanding contribution is the Government Information Quarterly, which is an international journal of information technology management, policies and practices in the United Kingdom and has seven publications related to search keywords and an impact factor of 5,098, which is a measure of the number of times an average article is cited in this journal. The journals with the second and third positions in the table are related to public health and the medical Internet. However, they have been considered because they integrate topics such as the Internet, marketing, new technologies, digital communication and public administration within their research subareas.

Table 3. Ranking of the most contributing sources (2011-2020)

Source	NP	PC	CO	JCR	SJR
Government Information Quarterly	7	5.8%	United Kingdom	5.098	2.121
BMC Public Health	5	4.1%	United Kingdom	2.521	1.198
Journal of Medical Internet Research	4	3.3%	Canada	5.03	1.446
Online Information review	3	2.5%	United Kingdom	1.805	0.624
Public Administration and Information technology	3	2.5%	United States	--	--
Sub National Democracy and Politics through social media	3	2.5%	United States	--	--
Profesional de la información	3	2.5%	Spain	1.58	0.698
Advances in Electronic Government Digital divide and regional development	2	1.7%	United Kingdom	--	--
Information Communication Society	2	1.7%	United Kingdom	4.559	2.806
International Journal of Public Administration in the Digital age	2	1.7%	United States	0.63	0.161

Journal of Social Marketing	2	1.7%	United Kingdom	1.94	0.599
Place Branding and Public diplomacy	2	1.7%	United Kingdom	1	0.295

Notes: NP= Number of publications; PC= Percentage of contribution; CO= Country of origin; JCR= Journal Citation Report; SJR= SCImago Journal Rank

The institutions (Table 3) that have more studies in the research area are the University of Hong Kong, with an annual average of 5 publications, being the most representative, followed by American universities such as the University of Florida, University of California, and the University of Georgia with an average of 4 publications per year. Among the universities with the highest scientific production, their world ranking was according to the QSp to demonstrate the high quality of the publications made annually. Some are in the top 40 worldwide, such as Stanford University (2), National University of Singapore (11), University of Hong Kong (25), University of California (28) and New York University (39).

Table 4. Number of Publication by Institutions (2011-2020)

Institutions	**NP**	**PC**	**CO**	**QSr**	**QSp**
University of Hong Kong	5	3.4%	China	83.8	25
University of Florida	4	2.7%	United States	47.7	167
University of California	4	2.7%	United States	82.6	28
University of Georgia	4	2.7%	United States	25	474
National University of Singapore	3	2.0%	Singapore	91.8	11
New York University	3	2.0%	United States	78.8	39
Stanford University	3	2.0%	United States	98.4	2
Syracuse University	3	2.0%	United States	-	581-590
Universidad de Almería	3	2.0%	Spain	-	-
University of California Irvine	3	2.0%	United States	41.3	219

Notes: NP= Number of publications; PC= Percentage of contribution; CO= Country of origin; QSr= Quacquarelli Symonds World University Rate; QSr= Quacquarelli Symonds World University Position

Finally, the countries with the highest academic publications were identified (Table 4). The United States ranked first with several publications at 34%, followed by Spain at 10.2% and the United Kingdom at 8.2%. There is a considerable difference between the leader's output and the rest of the countries, as shown in Table 4. It represents universities' and organisations' high interest in research in DCM and digital government.

Table 5. Number of publications by country (2011-2020)

Country	NP	PC
United States	50	34.0%
Spain	15	10.2%
United Kingdom	12	8.2%
China	12	8.2%
Australia	10	6.8%
Canada	7	4.8%
South Korea	5	3.4%
Germany	4	2.7%
Singapore	4	2.7%
Belgium	3	2.0%

Notes: NP= Number of publications; PC= Percentage of contribution

4. Conclusions

This bibliometric study was based on the review of 147 articles related to DCM and digital government in the WOS database between 2011 and 2020. The methodology used was mixed, where both qualitative and quantitative data were taken, and developing the study in two phases: first, the sample was taken and analysed in the VOSviewer tool, where the main keywords were identified (Figure 1), forming clusters that determined the relationship between each one (Figure 2). In the second phase, bibliometric indicators include the number of publications per research area, the sources that contribute most to scientific production, institutions, and the most productive countries.

The results showed in the first instance with the analysis of co-occurrence of the use of keywords such as "Social Media", "e-Government", and "Digital Government" within most of the literature consulted, in addition to others related as "Political Communication", "Media", "Engagement" and recurrent use of words such as "Twitter" or "Facebook" as primary communication channels for public entities (Figure 1 and 2). The implications arising from these findings will be helpful to any government to implement its e-government systems effectively and propagate them through the appropriate social media platforms for optimal dissemination and use by end-users.

On the other hand, in the frequency map (Figure 1) and the cluster of keywords (Figure 2), a gap was detected between terms related to the generation of content from digital channels for digital government, such as "Content Strategy" or "Content Marketing", there was no interrelationship. Its frequency was very low or almost zero, this being an opportunity to explore new avenues of research in content marketing. In addition, topics such as transparency of information come to perform an essential role in the data found due to the high incidence of issues related to social media.

In the last ten years, scientific production increased considerably regarding research topics, from 2018 to 2020 (Figure 3), the most significant number of scientific publications. The above is related to the increase in the adoption of technologies by public entities, potentiated by the pandemic in 2020 and the scientific community's interest in investigating these changes and topics. The number of publications per research area (Table 1) was mainly concentrated on Information Science, Business Economics and Government Law, thus evidencing that journals or sources of contribution also kept this relationship, adding one more research area such as Public Health.

In this way, journals such as BMC Public Health and the Journal of Medical Internet Research are part of the sources of the most significant contribution (Table 2). Those were taken into account by their research subareas related to the Internet, marketing, new technologies, digital communication and public administration. Besides, the study concluded that the prominent institutions (Table 4) that make the most contributions to scientific production are in countries such as the United Kingdom, the United States, Spain and China, with universities that are in the top 50 of the World Ranking (Table 3), proving the interest of these powers in doing research based on topics such as DCM and digital government.

Finally, the limitations of the study were given by the little relationship of terms in the question of DCM for digital government, and it was necessary to accomplish a more profound analysis with terms related to each one to determine the connection between them and mark a correlation that determines a significant relevance. The gap detected in the literature analysed related to content generation in digital e-government channels is an opportunity to explore new avenues of research in content marketing.

Chapter 3

Twitter analytical model

Analytical model to measure the effectiveness of content marketing on Twitter: the case of governorates in Colombia.

1. Introduction

Social networks have revolutionised how we communicate and share information, including governments around the world, eliminating geographical barriers and physical distances (Muñoz-Expósito, Oviedo-García and Castellanos-Verdugo, 2017). In recent years, the government's use of social media has become increasingly common and essential for political communication and transparency (Siebers, Gradus and Grotens, 2018; Joo, Lu and Lee, 2020). One of the main benefits of governments' use of social media is the ability to connect directly with citizens and listen to their concerns and needs (Eltantawy and Wiest, 2011; Feroz Khan et al., 2014; Zheng and Zheng, 2014). Governments use platforms such as Twitter, Facebook, and Instagram to share important news and updates and also adopt these platforms to collect comments and suggestions from the population (Criado, Sandoval-Almazan, and Gil-Garcia, 2013; Warren, Sulaiman, and Jaafar, 2014).

Another advantage of governments' use of social media is transparency (Gao and Lee, 2017). Social media allows governments to showcase their work and decisions in a more accessible and transparent way, which can help increase public trust in government (Ekman and Amnå, 2012; Siebers, Gradus and Grotens, 2018; Choi and Song, 2020). In addition, social media can also be a valuable tool for tackling misinformation and fake news, as governments can handle it to share verified information and debunk bottomless rumours (Bonson et al., 2012), which refer to unverified, unfounded, or false information that is spread without a solid basis or reliable evidence to support its veracity. These rumours usually spread quickly on social networks, generating confusion and misinformation.

Nonetheless, it is also important to note that governments' use of social media carries some challenges and concerns (Flynn, Nyhan and Reifler, 2017; Lazer et al., 2018). For instance, there are concerns about the privacy and security of information shared online, and there are also concerns about potential polarisation and hate speech online (Allcott, Gentzkow and Yu, 2019). Therefore, governments must ensure they are used responsibly and ethically to generate trust towards citizens (Siebers, Gradus and Grotens, 2018). The increasingly frequent use of Twitter as a marketing tool has led to a growing interest in measuring the effectiveness of content marketing on this platform.

Thus far, a comprehensive analytical model has yet to allow accurate and reliable measurement of the effectiveness of content marketing on Twitter. For this reason, this study aims to build an analytical model that determines which content characteristics improve the effectiveness of the Twitter accounts of the governorates of Colombia, taking as a measure of content effectiveness the Engagement Rate per Post (ERP) and identifying variability factors such as account followers, media types, posting time, use of hashtags and emojis and user interactions with the content (likes, shares, comments) (Table 6). In this way, it has been evidenced that emojis and hashtags are visual and textual ways of expressing emotions such as happiness, anger, sadness and love (LeCompte and Chen, 2017; Mayor and Bietti, 2021; Neel et al., 2023), which can impact the ERP of the content. Identifying and analysing these variables through machine learning algorithms is already possible, and it is great potential to measure incidents within the study. These algorithms are rapidly evolving, but there are three categories: Support vector machines (SVMs), Neural networks (Matsumoto, Yoshida and Kita, 2018; José, Juan Pedro Giudici and Luque, 2021), and Decision trees (Atif and Franzoni, 2022).

Table 6. Objectives and hypotheses related to each study variable.

Research Objective	*Hypothesis*	*Variables*	*Category*
Build a data analytics model that determines content characteristics to improve the effectiveness of the Twitter accounts of the governorates of Colombia.	H1: Twitter accounts with more followers generate a lower ERP compared to governorate accounts with fewer followers.	ERP (Engagement Rate per Post) versus number of followers	Quantitative
	H2: The contents that have some media (photo, video, link) have a higher ERP	ERP versus Media Type	Quantitative
	H3: The day and time of posting of the contents affect the ERP in the Twitter accounts of the governorates of Colombia	ERP versus Time/Date	Quantitative
	H4: The number of hashtags used in content on the Twitter accounts of the governorates affects the ERP	ERP versus number of hashtags	Quantitative
	H5: The use of elements such as emojis into the contents generates positive feelings associated with a high ERP	ERP versus emojis	Quantitative

In Colombia, 81% of the total population is an active user of social networks, around 41.8 million people, of which 4.3 million have an active Twitter account, 8.4% of the population and 12% of total Internet users. (Datareportal, 2022). According to the Ministry of Telecommunications in Colombia (MINTIC, 2021), social networks are the main communication channel between governors and citizens, far from using official websites. Very few citizens consult the websites, and the primary uses are to carry out transactional processes (Ramirez-Madrid et al., 2022). In addition, government programs, such as the "gobierno en linea" that has been implemented in Colombia over the past decade, are valued positively by citizens and are considered to influence the adoption of online government. However, these results indicate that there are still challenges regarding Internet access and public entities' use of web tools to communicate and interact with citizens (Torres et al., 2021).

2. Theoretical Background

2.1 Public Content Marketing (PCM) on Twitter

Social media marketing has become crucial to any content strategy (Pulizzi, 2012b). Social media has proven to be the most popular and effective way to engage with users and potential customers (Skoric et al., 2016). Organisations focus on engagement metrics to measure content success and effectiveness (Lei, Pratt, and Wang, 2016). These metrics may vary depending on the sector of the organisation. However, some general guidelines can be used to make measurements in the private and public sectors (Pulizzi, 2012a).

Public content marketing (PCM) serves to support the strategic priorities of a public or government organisation. It has become a powerful tool for teams managing digital channels (Henisa and Wilantika, 2021). In recent years, it has gained particular interest from the scientific community (Hollebeek and Macky, 2019; Holliman and Rowley, 2014). A favourable use allows the PCM to inform, educate and engage the public effectively and ultimately reflect the results of state entities' public programmes (Koob, 2021) and seeks to generate transparency, participation and collaboration (Chun et al., 2010).

Until recently, the PCM was considered a "good" in public management. This is because generating content through social networks was not a priority within public processes (Sweeney, 2019). Nowadays, the public sector understands the value PCM can have in reaching citizens, engaging and motivating them to make better decisions and generally undertaking positive actions or interactions with content posted on their social media accounts (Khan, 2017). Often the content created needs to be more structured with the absence or incorrect assignment of tags and mismanagement of PCM activities, such as content creation and dissemination (Salminen et al., 2019).

Moreover, there are many reasons why someone would want to participate on Twitter, but one of the most common is to convey a short and accurate opinion as a citizen (Ellison and Hardey, 2013; Choi and Song, 2020). For the government, the use of this social network has always been related to political content and electoral processes in order to influence citizens' decision-making (Gil de Zúñiga, Jung, and Valenzuela, 2012), and in this context, engagement is seen more like the participation of citizens in social or political issues (Skoric et al., 2016) and engagement rate can be seen in audience response to content, in the form of likes, comments, or actions (Schreiner et al., 2019). In addition, public organisations use Twitter to carry out advertising campaigns, public relations or as a means of customer service (Bonsón, Perea and Bednárová, 2019). Government-society communication on social media is limited and tends to be one-way only from the government, generating low interactivity rates with content (Henisa and Wilantika, 2021). The lack of a clear PCM strategy in public organisations means that a lot of the published content is not valuable to users. Bonsón and Ratkai (2013) argue that having a digital presence on social networks is not enough to achieve good engagement with citizens. Local governments must define better their digital communication strategies to maximise benefits (Bonsón, Perea and Bednárová, 2019) and to generate value with published content by better defining the profile of their audiences or communities (Pulizzi, 2012a).

2.2 El Engagement Rate per Post on Twitter

Measuring engagement on Twitter is a way to evaluate citizens' interaction with published content, have two-way communication and participate in policy formulation and decision-making (Henisa and Wilantika, 2021). User engagement can include actions such as retweets, likes, mentions or comments on a tweet (Vinerean and Opreana, 2021), and in terms of a deep engagement with the account, such as interacting with product or service content (Vivek, Beatty and Morgan, 2012), as well as participation in activities (Casaló, Flavián and Guinalíu, 2007; Wirtz et al., 2013) or co-creation activities in response to situations of complaints or claims (Bijmolt et al., 2010). Who is more willing to behave favourably by buying products or services (Bozkurt, Gligor and Gligor, 2021), thus generating several definitions highlighting their link with customer engagement (Jahn and Kunz, 2012). Measuring this type of interaction is essential for companies and public organisations that use Twitter as a PCM tool, as it provides deeper insights into citizens' behaviour with published content (Gordon, Baldwin-Philippi and Balestra, 2013).

The Engagement Rate per Post (ERP) on Twitter measures citizen participation in a particular publication (Mulyono et al., 2022). It is calculated as the number of interactions (likes, retweets, replies) that a publication receives, divided by the number of followers that the account that published the tweet has or by the reach the publication has, and multiplied by 100 to obtain a percentage.

$$ERP = \frac{\Sigma\ (favourites, retweets, mentions)}{total\ followers} *\ 100 \qquad (1)$$

There are some general metrics that are considered standards to measure citizen engagement on Twitter, as proposed by Bolson and Ratkai (2013) with the engagement calculation (Table 2). The metrics are divided into categories such as popularity, which can determine the audience reach of the content. Commitment is seen as the measure of comments per post. Virality measures the behaviour of retweets, and finally, the engagement rate considers all measures of the interactions described above.

Table 7. Calculation of citizens engagement

Metrics	Code	Measure	Details
Popularity	P1 P2 P3	Number of tweets/total tweets Total number of times favourite/total tweets (P2/number of followers)*1000	Average of favourite per tweet Total Popularity
Commitment	C1 C2 C3	Number of tweets commented/total tweets Total number of comments/total tweets (C2/number of followers)*1000	Average of comments per tweet Total Commitment
Virality	V1 V2 V3	Number of retweets/total tweets Total number of retweets/total tweets (V2/number of followers)*1000	Average of retweets per tweet Total Virality
Engagement		P3+C3+V3	

This calculation was complemented by Santoso, Rinjany and Bafadhal (2020), where it is proposed that the Stakeholder Engagement Index (SEI) which is not only a single measure and states that content can be analysed by labelling it into six categories, such as sharing information, calling for participation, communication with stakeholders, greetings and service information. From this categorisation, it is considered whether the content includes videos, links, photos and texts to determine its influence on the SEI calculation.

The percentage of ERP on Twitter considered high or low may vary depending on the sector or industry in which the account is located, as well as the goal of the particular account (Calderón-Monge and Ramírez-Hurtado, 2021). However, some general figures are considered standards for measuring ERP on Twitter. According to Sweeney (2019), there are ranges of the public sector industry that must be considered to classify the ERP. Between 0.09% and 0.33% is considered a high ERP and can be expected to reach 9 to 33 reactions per 1000 followers. Between 0.33% and 1% is considered very high and is expected to reach 33 to 100 reactions per 1000 followers. Therefore, It was identified that there are variables that have not been taken into account in the calculation of the ERP within the scientific community and that would potentially affect the effectiveness of the content in accounts of public organisations.

3. Methodology

In the search for a methodology that would adapt to the expected results of data mining, the two most used data mining processes were taken into account: CRISP-DM (Cross Industry Standard Process for Data Mining) and SEMMA (Sample, Explore, Modify, Model, and Assess). These two methodologies share the same essence, structuring the Data Mining project in phases that are interrelated with each other, turning it into an iterative and interactive process (Gomez et al., 2016), where their difference lies in the fact that CRISP-DM expands the process of understanding the business or object of study to transform it into a technical problem to be solved within practical scenarios (Jaramillo and Paz, 2015).

Thus, CRISP-DM provides a framework for organising and evaluating data mining tasks and is widely used in analytics and business intelligence (Schröer, Kruse and Gómez, 2021). This model generates an overview (Figure 1) of each phase needed in a data project with their respective tasks and outputs (Wirth and Hipp, 2000). As part of the methodology, the first phase, referring to the understanding of the business, was already presented in the literature review; the study will focus on the phases that involve the data (Figure 4).

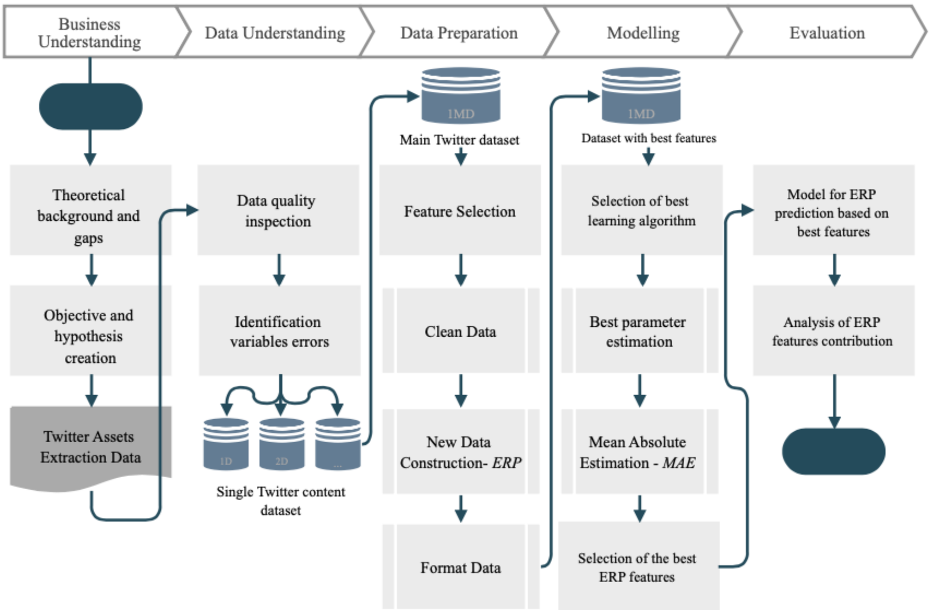


Figure 4. Methodological phases of CRISP-DM for a Twitter predictive model

3.1 Data understanding

Many of the studies consulted have taken European or Asian countries as case studies (Haro-de-Rosario, Sáez-Martín and del Carmen Caba-Pérez, 2016; Henisa and Wilantika, 2021; Bonsón and Ratkai, 2013; Mulyono et al., 2022; Santoso, Rinjany and Bafadhal, 2020; Bonsón, Perea and Bednárová, 2019), but few have been carried out in Latin America. For this reason, the governorates of Colombia were taken as the primary data source. They totalled thirty-two active Twitter accounts, from which data extraction was extracted in twelve months between January 2021 and January 2022 (Table 8). For the massive extraction of data, a web scraping tool was adopted, which allows automation of this process through algorithms to navigate the content of a page and extract the structured data for later analysis. Likewise, for data processing Python libraries such as pandas, numpy, and missingno helped to clean, standardise and construct new data; and for visualisation, the seaborn and matplotlib libraries were implemented.

Table 8. General data characteristics in Twitter datasets.

Activity	*Description*	*Result*
Data counting	Datasets consolidation	124.784 register
Types of values	Categorical, Numerical, Boolean	3
Variables	Followers, Media Type, Time, hashtags, characters, text, emojis	9

Thirty-two datasets of the information extracted from the accounts of the governorates were created, and each of the variables contained in the data sets denoted specific characteristics of the data (Saura, 2020). Most variables were numerical, which does not obey an intrinsic classification since they correspond to data such as the number of retweets, hashtags, mentions, likes, characters and emojis (Table 9). Similarly, categorical data refers to variables such as account name, URLs, text used in each post, language, tweet type, media type and media URL.

Table 9. Variable description in Twitter datasets

Variable	**Type of Variable**	**Description**
Account data		
Name account	Categorical	The official Twitter account name
URL	Categorical	Direct link to access the Twitter account
Followers	Numerical	Total number of followers of the account
Favourites /Likes	Numerical	Total number of likes on the account
Number of published posts	Numerical	Total number of posts in the account
Post data		
Tweet Id	Numerical	Number that identifies each post
Text	Categorical	Text used by post
Created At	Numerical	Post creation date
Favourites / likes	Numerical	Total number of favourites (likes) per post.

Retweets	Numerical	Number of times the post has been retweeted
Language	Categorical	Language used by post
Tweet Type	Categorical	The content distribution among three categories: tweets, retweets and replies.
Hashtags	Numerical	Total number of Hashtags used within the text.
Mentions	Numerical	Number of mentions by post
Media Type	Categorical	Media type available within the post: photo, video, gif.
Media URLs	Categorical	Link of the media type shared in the post.

No special coding schemes or schemes requiring specific detail were found. It should be noted that for some categorical variables, NaN values corresponding to empty data were found for a specific category type. Given the structure of the variables in the datasets, this is not constituted as missing data. In this phase, it was sought to know the volume, format, structure, quality, relationships, problems and possible solutions of the data and to be able to plan and design the data analysis project since it allowed to know the limitations and advantages of the data to be used (Saura, 2020).

3.2 Data processing

For the data processing phase, a new dataset was created from the fusion of the 32 datasets generated by each of the Twitter accounts of the Governors of Colombia. It was possible because all datasets share the same numerical and categorical variables, which underwent an initial inspection and cleaning process. Variables such as user ID, media type links, geolocations, language, user retweet id, and client name were removed from datasets. The variables of Tweet Id, Favourites, Retweets, Mentions, Text, Screen Name, Created at, Media Type, Tweet Type, Hashtags, and Mentions were prioritised. At this stage, the production of content from one account to another and the significant difference between the number of data collected from the accounts studied was preliminarily evidenced. Within the cleaning phase, duplicate data was eliminated through the Tweet ID to avoid biases in information processing (Sidorov et al., 2014). Similarly, the accounts' names were homogenised to prevent inconsistencies in the use of the datasets before the integration of the data since inconsistencies were found in the variable Name_account, possibly since the account's name can be changed in the social network, so it changed over time.

Furthermore, the preparation phase was one of the most important in the data analysis process, as clean and prepared data is essential to obtain accurate and reliable results in the following phases (Schröer, Kruse and Gómez, 2021). Once the datasets were cleaned and the variables that helped construct the model were selected, the new data began (Chapman et al., 2000). The ERP estimation was calculated following formula (1), and the numerical variables extracted from each dataset were taken to assess each of the hypotheses raised.

On the other hand, text analysis was performed using natural language processing (NLP) to extract valuable information from unstructured text data. The Text Blob algorithm was used, a well-known approach developed in the Natural Language Toolkit (NLTK) (Kunal et al., 2018). It involved tokenisation (separating the text into individual words) and eliminating irrelevant words. Moreover, for the preparation of categorical data and text processing of Twitter (Text and Hashtag), sentiment analysis was used, which evaluates the written content to determine the user's attitude or emotion towards the published content and classify it as positive, neutral and negative (Saura, Ribeiro-Soriano and Zegarra Saldaña, 2022). In addition, taking advantage of the use of programming strategies and machine learning algorithms, it was possible to process the hashtags, emojis and text to identify the emotions associated with the content shared by the accounts on Twitter, as follows:

- The bigrams and trigrams of the text associated with the tweet were constructed for the tokenisation process.
- A count of the number of hashtags and emojis within a text string was carried out to have a new autogenerated variable within the numerical type data set that later allows it to be evaluated against the ERP variable.
- Furthermore, the sentiment analysis based on the plain text was done through the Pysentimiento library (Python Toolkit for Sentiment

Analysis and SocialNLP tasks), which allows making pre-trained models based on neuronal networks to analyse the sentiment in text and to determine if it is positive, negative, or neutral (José, Juan Pedro Giudici and Luque, 2021) and the implication of using emojis to improve the content effectiveness (Neel et al., 2023).

3.3 Predictive model for ERP on Twitter

Character selection methods were used to obtain the optimal subset, as Khalid, Khalil and Nasreen (2014) suggested. The following steps were performed: five supervised learning algorithms were trained, and the performance metrics were compared. The subset that turns out to be the best will be the one used to create a prediction model of the ERP. To ensure the best results, a random-based model tuning process, available in the Scikit-learn pipeline module, was used to find the optimal configuration of each algorithm by adjusting it from variations of its hyperparameters (Lasso et al., 2020).

The algorithms tested were Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (Light GBM), Gradient Boosting Regressor, Random Forest Regressor and MLPRegressor. Cross-validation was done to measure how close the prediction was to the final results for each of the resulting models (Raschka, 2018), and the mean absolute error (MAE) was used as the evaluation metric. Thus, the best parameters were obtained for each algorithm applied to the dataset and its MAE. In addition, the original (not reduced) dataset was also evaluated with the same algorithms, and the MAE was obtained. Finally, the combination of the algorithm and the subset of data with the lowest MAE was selected.

The SHAP library for Python was used to analyse the importance of each feature in the proposed model, which allows estimating the SHAP values of each variable, which represent their impact on the final value of a prediction in a particular model from the analysis of the decrease in the performance of a model under different values of a variable (Lundberg and Lee, 2017). Dependency plots and a summary of SHAP values were used to represent the impact of each variable on the model output (Lundberg et al., 2020). It allowed the understanding of the contribution of each variable in the construction of the model, improving the calculation of the ERP.

4. Results

4.1 Data Understanding

The extraction of data from the thirty-two Twitter accounts of the governorates of Colombia yielded 124,784 (Table 8), of which 67,111 fulfilled the quality verification and were analysed. This distribution allows the study to evaluate the accounts individually and see the behaviour of the ERP with the most outstanding accounts. In the datasets, it was possible to demonstrate content generation production by some of the accounts, such as the Government of Antioquia, from which 5,412 records were obtained, in contrast to accounts such as the Government of Guaviare, where registrations fell by even 80%.

In the preliminary measurement of the ERP, there needed to be a value-added: the reach per post. Each account's number of followers in the data extraction period (January 2021- January 2022) was taken as a reference to mitigate the impact. Of the variables extracted (Table 9), five were data related to the account and eleven to the content; of these sixteenth variables, four were taken for the calculation of the ERP (Followers, Favourites, Retweets, Mentions) and the other variables were contrasted with this value to determine their incidence (Created at, Media Type, Tweet Type, Text, Hashtags).

4.2 Data processing

The data processing phase was essential to transform the data into valuable and accessible information. Multiple analysis techniques were used to explore the data and discover patterns and relationships between the variables to create the predictive model. Preliminarily, general statistical measurements were made to determine the behaviour of the ERP. The average ERP value for the dataset is 0.050 with a standard deviation of σ 0.213 (Table 10). The general mean of the ERP is between values of 0 to 4.9

and varies depending on the contrast with variables such as favourites 0.033, retweets 0.015, and mentions 0.0022.

Table 10. ERP statistical measures in Twitter

	engagement	**engagement_fav**	**engagement_rt**	**engagement_ment**
count	67111	67111	67111	67111
mean	0.050832	0.033034	0.015542	0.002256
std	0.2131118	0.145174	0.069712	0.011752
min	0	0	0	0
25%	0.001763	0	0	0
50%	0.012987	0.007751	0.002218	0
75%	0.036942	0.02469	0.009934	0.001384
max	4.926075	3.299731	1.899027	0.564972

The number of followers on a Twitter account can significantly impact engagement. It was evidenced that as the number of followers increased, the ERP of the publications did not increase significantly, confirming the H1. Accounts with more than 100,000 followers have an average ERP value between 0.001 and 0.05 and are considered low in public sector industry ranks (Figure 5). Besides, it is evident that the accounts with the best ERP have an average of followers between 50,000 and 100,000, and their content production is less than 5000 posts per year. The outliers showed two accounts, one in which the number of followers was meagre, and the ERP soared above average to 4.92. Further, another account whose number of followers was above 400,000 showed an ERP of 0.001.

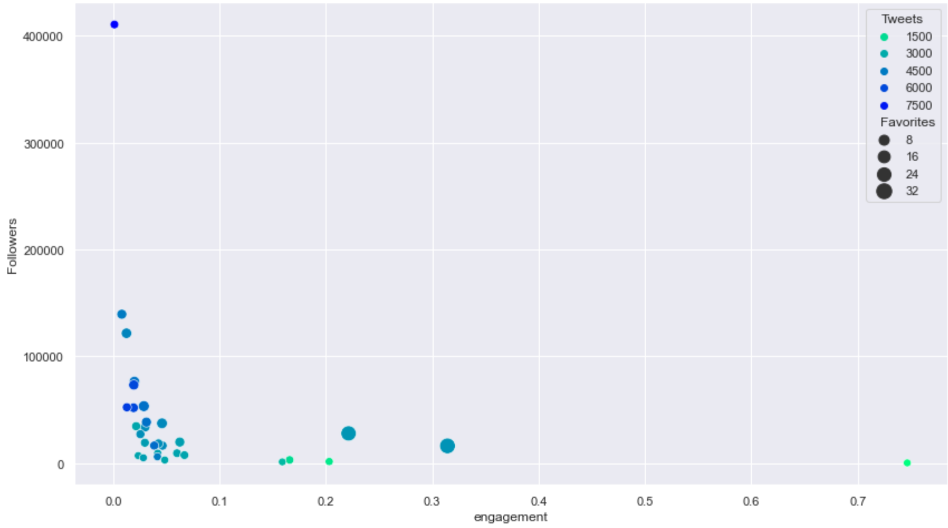


Figure 5. Relationship between the number of followers and the ERP

The overall content measurement showed that high production is not an improvement factor for ERP. On the contrary, it is evident that the accounts with the best ERP have an average of favourites above 30, and the generation of content ranges between 3000 and 5000 posts per year. Several tweets above 5000 per year generated an average interactivity of 10 likes per tweet and an ERP below 0.15, which happens for accounts with the highest number of followers (Figure 6).

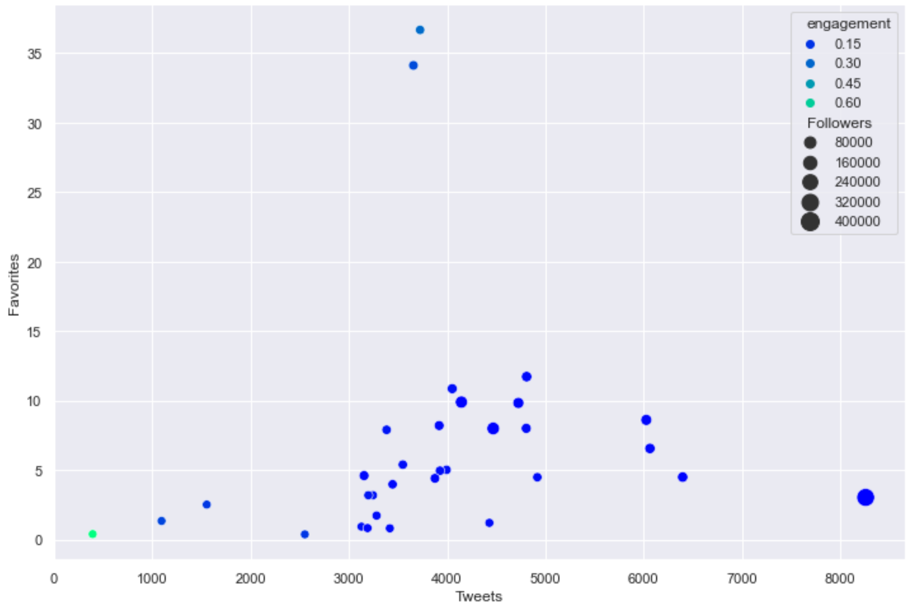


Figure 6. Relationship between the number of tweets and the ERP

On the other hand, it was evidenced that the media type (MT) used in the content significantly impacted ERP engagement (Figure 7). The most used MT in content is MT-video, with an average ERP value of 0.08 regarding your own content (tweet) and an average of 0.01 when sharing content from other accounts. Similarly, using MT-photo generated ERP values of 0.078 and 0.03, with the generation of own content having a more significant impact. The type of Tweet (Tw-Type) is also a variable that influences the ERP. The ERP rate increases if the content is original to the Tweet account (TW) and contains MT-photo or MT-video. Besides, shared content from other Retweet accounts (r-TW) generates that this rate lowers its performance. It is evident that for TW content, MT-video is the one that registers the best ERP, confirming H2.

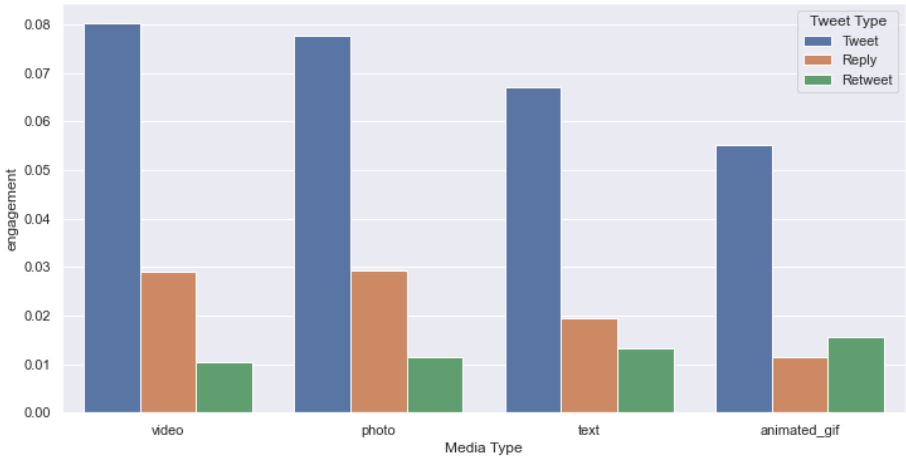


Figure 7. Relationship between the use of media type and ERP

The relationship between posting time (PT) and ERP are shown in Figure 8, where variables of posting time in a day (PT-*day*) and time of posting in an hour (PT-*hour*) were measured. For this, a heat map allowed contrasting values and established an ERP measurement scale between 0.00 and 1.00. The green colour represented lower values and the red values higher. The initial dataset had data with different time zones due to the different data extraction techniques. It required standardising PT-hour variables to Colombia Standard Time (GMT-5). The highest value recorded for the ERP was 0.096, with a PT-day Monday and TP-hour 8:00, followed by PT-day Friday and PT-hour 8:00 with an ERP of 0.086. On the other hand, the lowest values recorded were PT-day Thursday and PT-hour 14:00 of 0.033. Identifying a general trend in the heatmap was impossible due to the difference in content handling that Twitter accounts have (H3). As a result, the account of the Government of Cesar was analysed since it had the highest ERP, 3.70, to determine users' behaviour with the content and the hours of posting. The highest value recorded for the ERP was TP-day Thursday and TP-hour 6:00 of 3.70, followed by TP-day Friday and TP-hour 5:00 with an ERP of 3.22 (Annex 1)

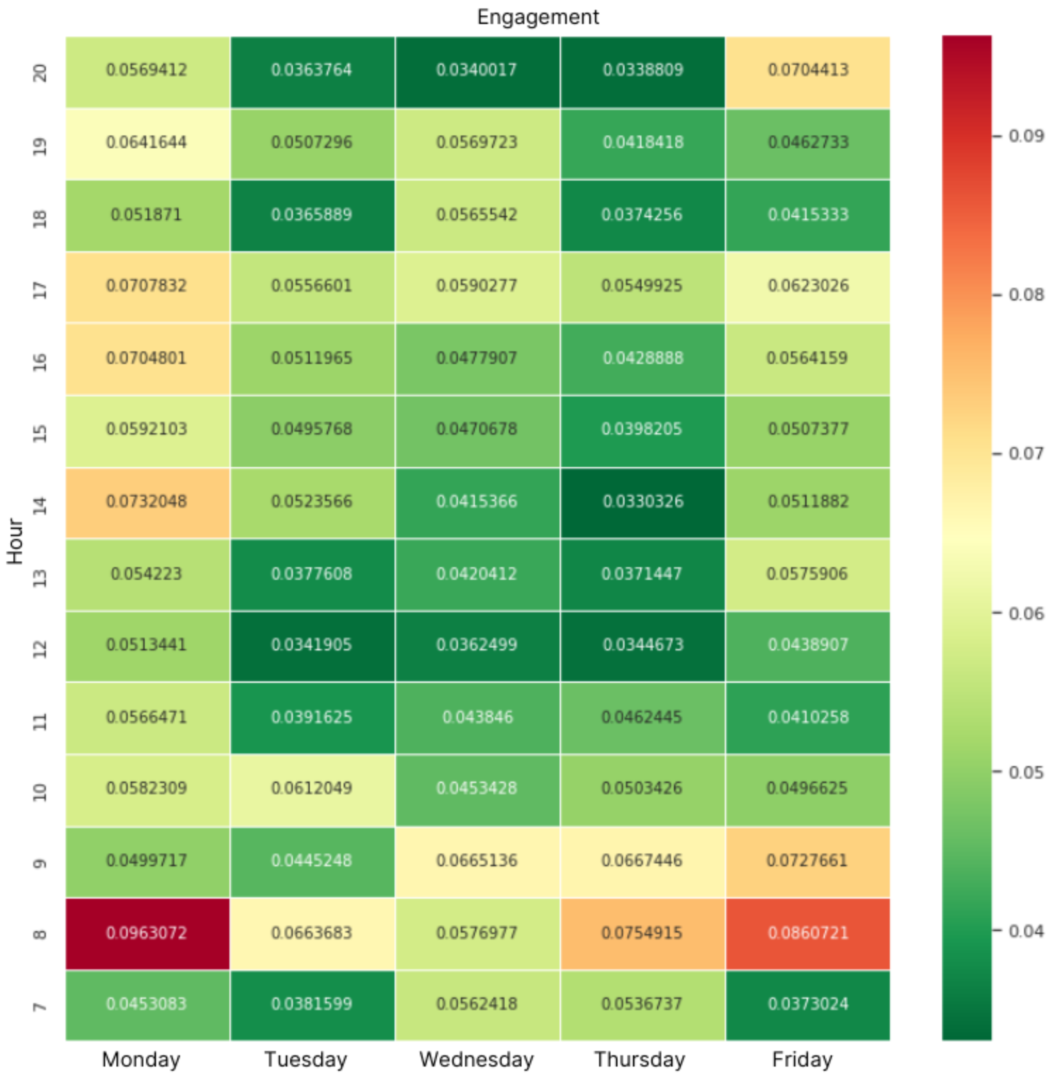


Figure 8. Relationship between TP-*hour*/TP-*day* and ERP

Using hashtags (HT) in Twitter content can significantly impact account engagement, as was proposed in H4. HTs are a way to tag content and make it easier for the audience to find. A finding of the study was that the use of

HT affects positively. The ERP reaches a value close to 5 if there is from 0 to 2 HT within the content. However, the number of HT within the > 5 content generates ERP values below 1 and, in some cases, close to 0. In addition, it was found that the contents that did not use HT had MT-video or MT-photo increased the ERP (Figure 9).

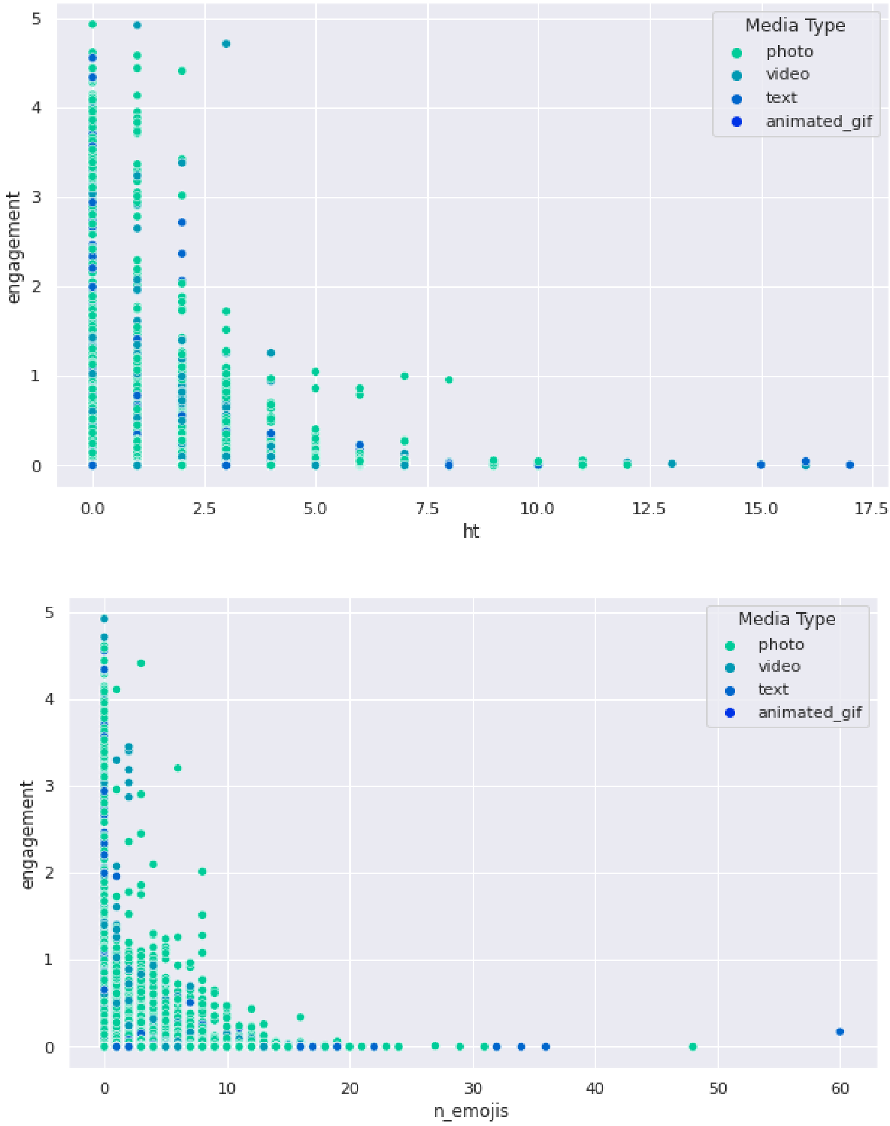


Figure 9. Relationship between hashtags/emojis and ERP

Similarly, using emojis (EM) within the Twitter content affected the ERP's behaviour (H5). When its use was from 0 to 3, the ERP reached a value between 3 and 5 (Figure 9). On the contrary, the number of EM within the >3 content generates ERP values below 1 and, in some cases, close to 0. In addition, as well as the use of HT, it was found that content that did not use EM and had MT-video or MT-photo caused the ERP to increase.

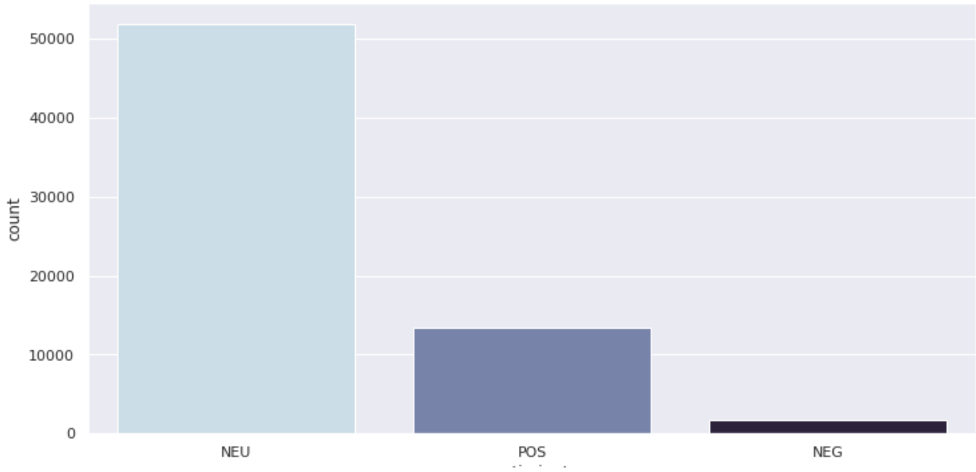


Figure 10. Content sentiment analysis in Twitter

The sentiment analysis required a transformation and standardisation of the text (Figure 10). For this, unigrams ("rt", "today", "Governor"), bigrams ("new cases", "Huila grows", economic reactivation") and trigrams ("rt @fndcol:", "rt @luismonsalvo:") were created to determine which were the most common words within the content. Likewise, lowercase and accent standardisation were eliminated. As a final result, it was evidenced that mostly (>50,000 posts) the content has a neutral sentiment followed by a moderate amount of positive sentiment (12,000 posts) and low content with negative sentiment (< 2,000 posts). It is considered that using elements such as EM or HT does not necessarily generate positive feelings that improve the ERP.

4.3 Predictive model for content variables on Twitter

The construction of the model was based on the correlation of characteristics and importance values given by the Light GBM Algorithm. The resulting MAE was 0.0565, with a mean square error (MSE) of 0.0441 and a root mean square error (RMSE) of 0.2103. These values were obtained after running the pipeline to find the best configuration of the regression algorithms (Table 11). According to the evaluation results of the regression algorithms for the correlation of variables, the set of variables that most affect the engagement per post are: the tweet type, the use of emojis, the dates (day, time), the type of media (photo, video, gif), the sentiment associated with the post and the emotions.

Table 11. Pipeline results to find the best regression algorithms

	Model	**MAE**	**MSE**	**RMSE**
XG Boost	Extreme Gradient Boosting	0.0574	0.0444	0.2107
LightGBM	Light Gradient Boosting Machine	0.0565	0.0441	0.2103
GBR	Gradient Boosting Regressor	0.0571	0.0444	0.2108
RFR	Random Forest Regressor	0.0566	0.0442	0.2102
MLP	MLP Regressor	0.0644	0.0446	0.2112

The values obtained from the algorithm training were summarised in the SHAP values for the characteristics according to the ERP (Figure 11). The range of values for each feature is represented in a colour gradient, where red represents its highest value and blue is the lowest. It was required that for the optimisation of the model and its visualisation, categorical variables were coded as binary variables.

Figure 11 is represented by SHAP values, as the violin graph represents the value of each variable on its own scale. The impact on the model's output is linked to the base value, and its axis shows how the prediction varies above or below this value on a scale of -0.08 to 0.08. When analysing the three main features of incidence in the ERP, Tweet type_tweet has the highest in the model, up to 0.050. In this way, its maximum value (red) is considered as TW (Tweet) and its minimum value (blue) is r-TW (retweet), thus considering that the TW-type content increases the expected incidence. A high number of emojis (EM) lower the predicted incidence, and the time (TP-hour) allows determining that content published earlier in the day increases the expected incidence.

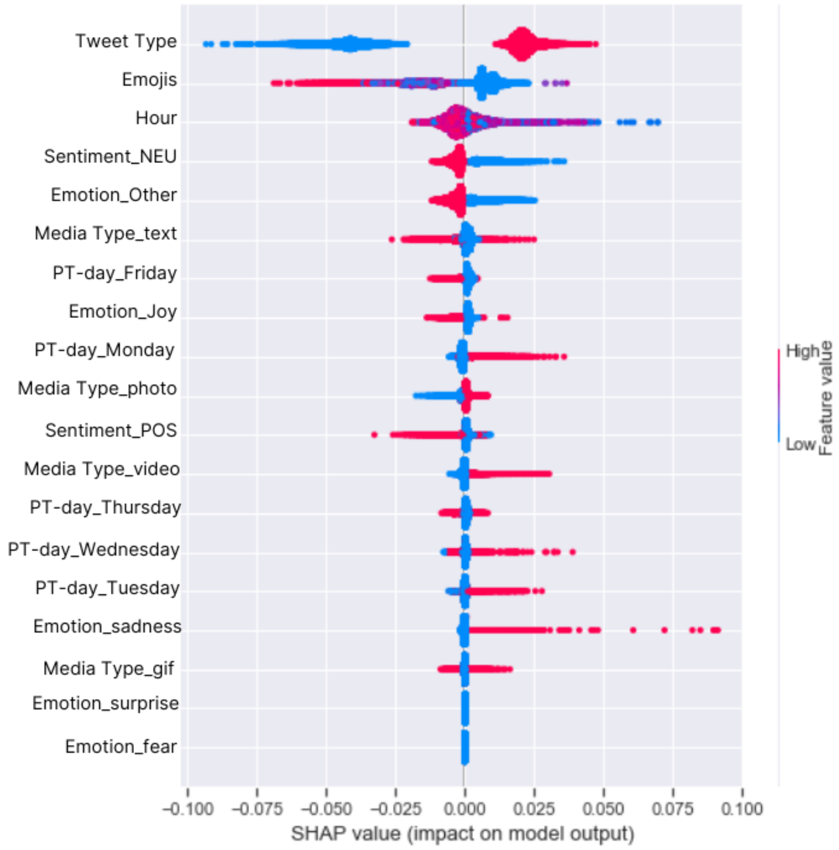


Figure 11. Summary of SHAP values for Twitter model

5. Discussion

With Twitter, governorates have improved communication with their citizens, allowing a more direct and personalised interaction to build trust and generate a sense of transparency and accountability, as stated by Siebers, Gradus and Grotens (2018) and Skoric et al. (2016). It has allowed social networks also to become a tool to mitigate misinformation and fake news, in which the government can generate official digital content (Bonsón et al., 2012). If the government uses social media effectively, it increases its reach and improves its public image, as evidenced by some of the accounts studied with more robust digital strategies. However, the need for governments to strengthen their digital communication strategies to maximise their benefits continues to be evidenced (Bonsón, Perea and Bednárová, 2019) through valuable content and ERP improvements that allow better interactivity with users.

According to the research objective, the Twitter accounts of the governorates of Colombia were studied in their entirety, thus covering 100% of the national territory. Preliminarily, the initial data exploration showed a big difference between significant departments that have extensive content production and departments that are farther away with less activity on Twitter.

The model allowed to analyse the different variables that had not previously been taken into account in the measurement of the ERP and thus determine the effectiveness of the CCM. The previously proposed models, such as that of Bonsón and Ratkai (2013), exempted variables such as the type of content, the use of media types, and the time of posting, while in that of Henisa and Wilantika (2021), there is no evidence of measurements with hashtags and emojis. The results showed that Twitter accounts with many followers (>100,000) presented shallow ERP values, close to 0. Three reasons were identified why this could happen: first, it is because the distribution of the content does not necessarily occur for its entirety of followers, second the quality of the content, users do not take it as valuable; and finally, the behaviour of users is not active within the social network. Likewise, the amount of tweets generated per account is not a determining factor for improving the ERP. It means accounts with more than 8,000 posts per year did not show improvement in this measurement. In addition, this significant difference between accounts for the number of followers and content generation explains the high coefficient of variation (4.9) that occurred in the measurement of the ERP, which was expected due to this variety of conditions.

Further, it was demonstrated in this study as in previous studies (Haro-de-Rosario, Sáez-Martín and del Carmen Caba-Pérez, 2016; Henisa and Wilantika, 2021; Bonsón and Ratkai, 2013; Mulyono et al., 2022; Santoso, Rinjany and Bafadhal, 2020; Bonsón, Perea and Bednárová, 2019), that TM remains an incidence variable in PRD. MT-video and MT-photo showed the highest ERP rates, as opposed to content with only text or gif. One of the most significant findings in this measurement was to consider the Tw-Type, contents of the account (TW) generate better ERP. In contrast, content such as replies or retweets (r-TW) lowers the performance of the ERP by more than 60%, evidencing in preliminary studies the neglect of this variable.

The posting time (TP) showed that the highest ERP values are TP-day Monday and TP-hour 8:00 and TP-day Thursday and TP-hour 8:00 when the account presents more significant activity in generating content and interactivity with users. This time differs precisely from that proposed by Henisa and Wilantika (2021). The days that affect significantly are Wednesday, Saturday, and Sunday. Meanwhile, the post times that influence citizen engagement the most are TP-hour 06:00-09:59 and 18:00-05:59 (Twitter). It shows that PT is a variable that does not have consistent behaviour. Even within the same study with the 32 accounts, it was not possible to build ranges of better interactivity but remarkable results of ERP values. The ERP reaches significantly high values even if they are not used, or the content has at least one. On the other hand, HT generates low ERP values in the content when its use is greater than 3, regardless of the use of MT-video or MT-photo.

5.1 Limitations and future research

One of the study's main limitations was not having scope data per post, which refers to the total number of people who have viewed a specific post. Instead, we have to use each account's total number of followers. This could improve the refinement of the predictive model since it would allow us to have more accurate values for training and adjust the model parameters. For future studies, this model should be taken with scope values to replicate the research in different industries, not only in content for government entities.

Another research limitation was to capture data longer than two years to determine the behaviours of the accounts and thus be able to feed the model with more training data. Preliminarily, it was identified that comments as a measurement variable in the ERP could be counterproductive because if there is a greater number of comments than likes, it is perceived as content with little acceptance among users. For future studies, this behaviour should be further studied through a machine learning model.

6. Conclusion

The evaluation of the effectiveness of these contents is based on formulas that are transversal to private industries, which restricts, to some extent, the final objective of the contents. The literature consulted has evidenced the measurement variables of the traditional Engagement Rate per Post (ERP), which includes the metrics of interactivity (likes, shares, retweets and comments) against the reach of the account (followers). It generates unique ratios of content effectiveness, but they could be more flexible. To address this limitation, we sought a methodology that allowed organising the data project and adapting to a Content Marketing (CM) process to build an analytical model determining which content characteristics improve effectiveness.

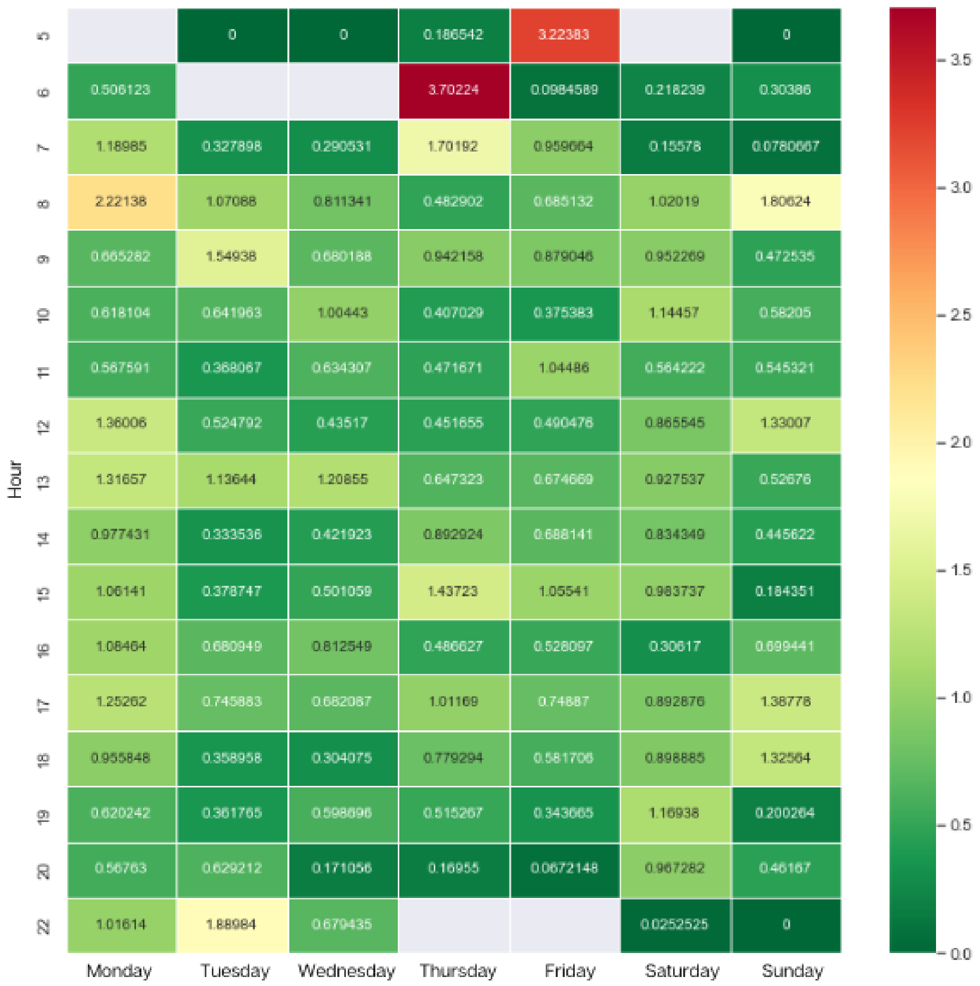
The CRISP-DM methodology provided a framework for organising and prioritising data mining tasks. The first phases allowed us to understand the gap in the literature versus the effectiveness of the content, define the objective and hypotheses of the study, and determine the technologies for the extraction and evaluation of data quality. The subsequent phases focused on processing and analysing data using machine learning algorithms to identify patterns in the variables that would determine which of them most influence the effectiveness of the ERP.

Based on these results and the analysis obtained, it was possible to reach the objective of building an analytical model that would determine which characteristics of the content improve the effectiveness of the Twitter accounts. The training model analysed variables such as tweet type, number of followers, type of content (photos, videos, gif), hours of publication, use of hashtags and emojis, and user interactions with content (likes, shares and comments). It was identified that, although these variables directly influence the ERP. Likewise, the most influential variable to improve the ERP was the content generated by the accounts as their own (Tweet), since when retweets were made from other accounts, the ERP decreased by more than 80%, generating little interactivity with users. Likewise, it was found that emojis within the content for this type of account directly affect the ERP, which requires special attention from government organisations. This finding allowed us to establish that there are factors that have yet to be taken into account to measure the ERP and that have a critical load on users' interactivity with the content.

This study has contributed to the literature by addressing the need to develop a comprehensive analytical model to measure the effectiveness not only of public content marketing (PCM) on Twitter, rather it was determined to be highly replicable in different industries or companies. The results displayed the variables that most influence the effectiveness of the content, suggesting that more than traditional metrics of interactivity and reach may be needed to accurately and reliably measure the success of content marketing in this type of account. Instead, the content generated by their accounts and the use of emojis is presented as critical factors in improving engagement with users. The model gives professionals an effective way to measure and understand the ERP in-depth, allowing them to make more informed decisions and optimise their content strategies to achieve better results. Also, the findings of this study can serve as a base for future research that seeks to improve the measurement of the effectiveness of content marketing on Twitter and, in particular, for the construction of more complex analytical models adapted to the different contexts and objectives of the accounts.

ANNEX

Annex 1. Relationship between TP-hour/TP-day and the ERP of the Government of Cesar



Chapter 4

Facebook and Instagram analytical model

Measuring the effectiveness of content marketing on Facebook and Instagram: an analytical model for Governorates in Colombia

1. Introduction

Governments have a responsibility to be accountable to citizens and society as a whole since their main task is to protect the interests of the general public. With the appearance of the Internet and social networks, it has become possible to publish a large amount of information in an affordable and accessible way (Cegarra-Navarro, Garcia-Perez and Moreno-Cegarra, 2014; Bonsón, Royo and Ratkai, 2015). Governments began to use Facebook and Instagram as communication channels several years ago; it could be affirmed that since these platforms were created (MacKay et al., 2022), they saw them as channels of communication and amplification of information. Facebook and Instagram have a global user base of 2,989 million and 1,444 million, respectively (Datareportal, 2023). This massive reach gives governments an unprecedented platform to disseminate information and reach a broad audience in real-time (Thackeray et al., 2012; Gruzd, Lannigan and Quigley, 2018; de Jong, Neulen and Jansma, 2019). Through regular posts, announcements, and multimedia content, governments can deliver key messages about public policies, government programs, and relevant events on an unprecedented scale (Pang et al., 2021).

One of the most powerful features of Facebook and Instagram is the ability to establish two-way communication with citizens (Al Aufa, Sulistiadi and Djawas, 2020). Governments can engage citizens in meaningful dialogues through comments, private messages, and interactive polls. According to Vragov and Kumar (2013), the active use of these technologies not only encourages greater citizen participation but also allows governments to understand better their population's needs, concerns, and opinions, creating a relationship of trust and openness (Enria et al., 2020). Each government has diverse audiences with specific needs and priorities, such as promoting open government (Wirtz, Daiser and Mermann, 2017). Thus, Facebook and Instagram offer advanced demographic targeting tools, allowing governments to tailor their message and content to specific audiences. This ability to personalise helps ensure that information is relevant and reaches the proper citizens, thus maximising the impact of communication (Gruzd, Lannigan and Quigley, 2018). For his part, Linders (2012) presents a typology for electronic citizen participation that includes: citizen supply, government as a platform, and self-management government. These categories reflect the models of citizen participation in the era of social networks: citizen-to-government (C2G), government-to-citizen (G2C) and citizen-to-citizen (C2C), respectively. C2G mainly focuses on consultation and idea generation, where citizens can share their opinions with the government. G2C primarily focuses on information and momentum, where citizens are empowered with data to make informed decisions. Finally, C2C is about citizen self-organisation.

Similarly, managing social networks for governments has become a tool to identify false news (Bonson et al., 2012) and even deny conspiracy campaigns (Song and Lee, 2015). In emergencies, breaking news or critical government announcements, speed of communication is critical (Rustad and Sæbø, 2013). Facebook and Instagram provide governments with the ability to transmit information in a prompt and timely manner. In this way, the type of content that is published begins to vary depending on the needs and objectives to transmit information, as well as the variety of types of content that is allowed to be generated by each of the platforms to improve the effectiveness of communication (Koob, 2021).

Given the use of these social networks and the amount of content that circulates through them, the measurement of effectiveness is an identified need (Bonson, Royo and Ratkai, 2015; Henisa and Wilantika, 2021; Hollebeek and Macky, 2019; Santoso, Rinjany and Bafadhal, 2020; Pang et al., 2021) to develop an analytical model that allows the identification of factors and characteristics that improve or affect this effectiveness in the content on Facebook and Instagram, based on data from the Colombian governorates. For this, the engagement rate per post (ERP) was used as a measure of content effectiveness, which gives us user interactivity rates with the content, and variable factors were analysed, such as the number of followers of the account, the types of media used, the time of publication, the use of hashtags and emojis, as well as user interactions with the content. These aspects are detailed in Table 1.

In the Colombian context, Facebook and Instagram have achieved a significant impact, providing unique opportunities for communication, social interaction and the dissemination of information. Facebook and WhatsApp have gained unprecedented popularity in Colombia, with approximately 90% of Internet users having an account on at least one of these platforms (Arroyave Cabrera, 2020). Data shows steady growth in Facebook adoption in the country, with 32 million users in 2019 and an estimated 35.15 million in 2022 (Kemp, 2022). These statistics reveal the broad influence of Facebook in Colombian society and its fundamental role as a channel of communication and social interaction. On the other hand, Instagram has yet to reach the same magnitude as Facebook regarding users in Colombia, but it has experienced constant growth in recent years. Approximately 20% of Colombians use Instagram, but most notably, more than 60% of users are between 18 and 34 (Mencia, 2020), making it an attractive platform to reach a young audience. This opens up a range of opportunities for governments and other entities interested in communicating and connecting with the youngest population in the country. Social media users in Colombia increased by 2.8 million (+7.2 per cent) between 2021 and 2022, and the number of social media users in Colombia at the beginning of 2022 was equivalent to 81.3 per cent of the total population (Kemp, 2022).

2. Theoretical Background

Government social media management requires a sector-specific approach (Falco and Kleinhans, 2018). These tactics that work for private companies can sometimes be applied in different ways in social networks in the public sector (Wirtz, Daiser and Mermann, 2017). For this reason, in a private sector context, the administration of social networks such as Facebook and Instagram may have more freedom regarding content, engagement and regulations. However, this tactic may apply to us in the same way in the social networks of governments (Eom, Hwang and Kim, 2018). Many of the challenges in managing social networks in the public sector are due to the connection that must be made between the government and citizens (Haro-de-Rosario, Sáez-Martín and del Carmen Caba-Pérez, 2016). This connection is given by strategies such as public content marketing (PCM) (Henisa and Wilantika, 2021), which refers to creating and sharing relevant, valuable, and consistent content through digital media to attract and retain an audience in specific segments. For this reason, the measurement of the effectiveness of the content plays an essential role since it allows for calculating the participation of citizens (Choi and Song, 2020) with a specific publication on social networks and represents the level of commitment of this audience with shared content (Warren, Sulaiman and Jaafar, 2014).

The engagement rate per post (ERP) is a valuable tool for marketers and social media managers, as it allows them to evaluate and compare the performance of different posts (Bonson, Royo, and Ratkai, 2015) and adjust their strategies accordingly function of the results (Ekman and Amnå, 2012). It can also be used to perform trend analysis (Falco and Kleinhans, 2018) and determine what type of content improves effectiveness (Santoso, Rinjany and Bafadhal, 2020) among G2C. The ERP on Facebook is calculated by dividing the number of interactions (likes, comments and shares) of a publication by the total reach of that publication and multiplying the result by 100 to obtain a percentage (formule 2). However, on Instagram, shares are not considered since the platform does not allow it, but there are metrics for sending and saving.

$$ERP = \frac{\Sigma\ (likes, comments, shares)}{total\ followers} * 100 \qquad (2)$$

For this reason, just taking these metrics can make ERM a fancy measure. In order to achieve a more appropriate content measurement for Facebook and Instagram, what is proposed by Bonsón, Royo and Ratkai (2015) and what is proposed by Santoso, Rinjany and Bafadhal (2020) are taken as a reference to adapt it to the new metrics of these platforms (Table 12).

Table 12. ERP Calculation on Facebook and Instagram

Metrics	Code	Measure	Details
Reach	R1	Number of post “liked”/total post	Average of like per post
	R2	Total number like or reactions/total post	
	R3	(R2/number of followers)*100	Total Reach
Interaction	I1	Number of post commented/total post	Average of comments per post
	I2	Total number of comments/total post	
	I3	(I2/number of followers)*100	Total interaction
Amplification	A1	Number of shares/total post	Average of Shares per post
	A2	Total number of sharess/total post	
	A3	(A2/number of followers)*100	Total Amplification
ERP		P3+I3+A3	

In some situations, an account with a large number of followers can generate a low ERP if the interaction and participation of those followers are low. This could be due to several factors, such as the quality of the content (Lev-On and Steinfeld, 2015), the relevance to the audience, and the communication strategy (Guillamón et al., 2016) since Instagram users exhibit a significantly higher problematic usage behaviour compared to Facebook users. (Limniou, Ascroft and McLean, 2021). If followers are not engaged or motivated to interact with posts, ERP may be low or fluctuating, even with many followers (Lappas et al., 2018).

Hypothesis 1 (H1): The greater the number of followers on Facebook and Instagram accounts, an ERP with low values is generated.

On the other hand, the type of media used in the content, be it a photo, video or sidecar, can significantly impact followers' engagement (Putranto et al., 2022). Visual content tends to be more engaging and effectively captures the audience's attention (Gruzd, Lannigan and Quigley, 2018). Images and videos are eye-catching and convey messages more quickly and effectively than textual content (Maiano et al., 2021). Visual media also helps to generate emotions, arouse the followers' interest and motivate them to interact with the publication through likes, comments or shares (David, 2010). In some cases, visual content can be more entertaining, informative, or inspiring, resulting in higher engagement and engagement from followers (Haenlein et al., 2020).

Additionally, using relevant and popular hashtags in Facebook and Instagram posts can help categorise and tag content (Brodie et al., 2011; Saxton et al., 2015), making it easier for interested users to discover issues related to government or issues of citizen interest (Hemphill, Culotta and Heston, 2013). Hashtags allow publications to be more easily found in searches and on the results pages of the platforms' internal search engines (Mas-Manchón and Guerrero-Solé, 2019).

Hypothesis 2 (H2): Using media type (Photo, Video, Link) in the content generates high ERP values.

Hypothesis 4 (H4): Using hashtags in Facebook and Instagram content leads to higher ERP values

Another variable that potentially affects the ERP is choosing the right day and time to publish the content since the reach can be maximised and more significant interaction generated by followers (Cheng, Danescu-Niculescu-Mizil and Leskovec, 2014). Some studies (Kapetanaki, Bertele and Brennan, 2017; Cuevas-Molano, Matosas-López and Bernal-Bravo, 2021; Eslami, Ghasemaghaei and Hassanein, 2021) and data analysis have shown that there are patterns and trends regarding the days and the hours when users are most active on social networks.

Hypothesis 3 (H3): The day and time of content posting allow to identify the highest ERP values.

Emojis allow for adding an emotional and personal component to content, generating a more direct and empathetic connection with the audience (Boutet et al., 2021; Neel et al., 2023). Emojis that convey positive feelings, such as smiles, hearts, or thumbs up, can generate a positive emotional response in followers (Ayvaz and Shiha, 2017). This can increase engagement with the content, such as the number of likes, comments, and shares.

Emojis can also help convey the tone and intent behind a message (Erle et al., 2021). In an online environment where written communication can be misperceived or lack emotional nuance (Padgett, Moffitt and Grieve, 2021), emojis can help clarify meaning (Fugate and Franco, 2021) and generate a more positive and supportive response favourable by the followers.

Hypothesis 5 (H5): Using emojis in the content generates higher ERP values associated with positive sentiments.

3. Methodology

In the early 2000s, data research techniques or data analysis emerged when the term KDD (Knowledge Discovery in Databases) was used to refer to the (broad) concept of finding knowledge in data (Rodriguez Leon and Garcia Lorenzo, 2016). Thus, to standardise data mining processes, two primary methodologies were developed, CRISP-DM (Cross Industry Standard Process for Data Mining) and SEMMA (Sample, Explore, Modify, Model, and Assess) (Jaramillo amd Paz, 2015). Both describe the tasks to be carried out in each phase, assigning specific tasks and defining the desired result after each phase. (Wirth and Chrysler, 2000).

In their study, Azevedo and Santos (2008) analysed the two implementations. They concluded that, although there is an apparent similarity between them, CRISP-DM is a complete methodology because it considers applying the results to the business environment. Therefore, it was the one that was widely adopted in many industries (Peker and Kart, 2023). According to surveys conducted by Saltz (2020), CRISP-DM has consistently been the most popular methodology for analytics, data mining, and data science projects through 2020. The CRISP-DM methodology provides an iterative approach, including frequent opportunities to assess the project's progress concerning its original objectives (Martinez-Plumed et al., 2020). It also encourages data miners to focus on business objectives, ensuring that project results provide tangible benefits to the organisation (Schröer, Kruse and Gómez, 2021). In addition, the methodology is both technology and problem neutral, allowing the use of any data analysis and processing software (Rotondo and Quilligan, 2020).

Based on this, a model was created for the study. Figure 12 shows the stages necessary for the study and their respective tasks. As part of the methodology, the first phase, which involves understanding the business, has already been presented in the literature review, so the study will focus on the phases that involve understanding and preparing the data, in addition to the modelling phase for the construction of the predictive model.

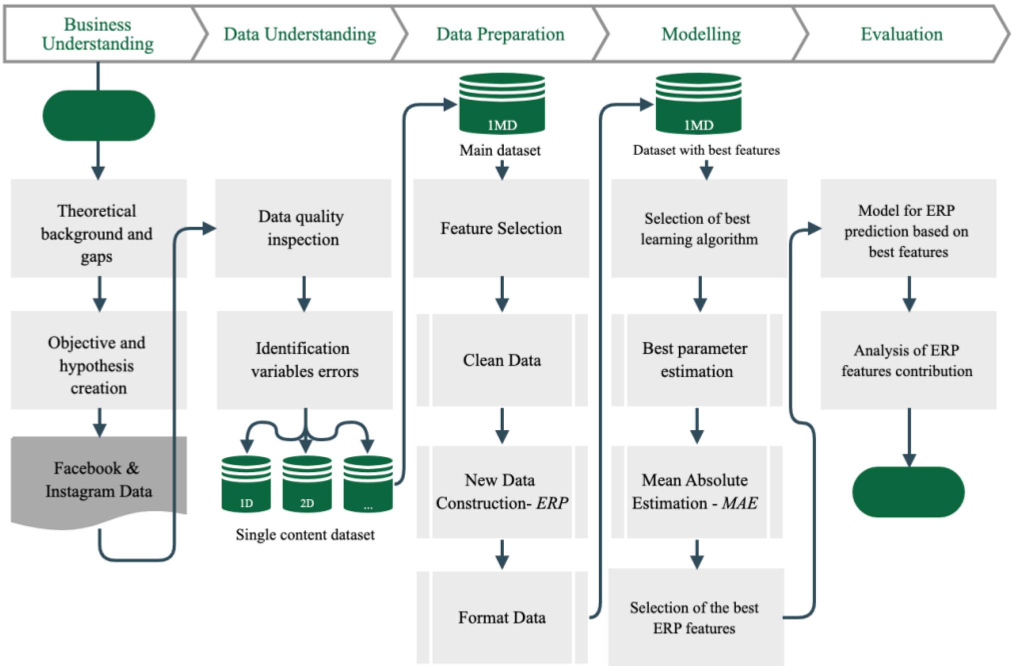


Figure 12. CRISP-DM process for Facebook and Instagram predictive model

3.1 Data understanding

The second stage of the CRISP-DM methodology is the data understanding phase. The goal of this phase was to gain a general understanding of the data to be worked with. Several tasks were performed at this stage, such as gathering information about the data, its origin, format, content, and quality, among other things. To gain a general understanding of the data's distribution, patterns, and trends, graphical and statistical tools were used. In addition, data quality problems such as missing, duplicate or inconsistent values were identified, and the variables relevant to the analysis and how they relate to the project objective were identified. The construction of the general characteristics of the data was also carried out where activities such as data counting, the types of values and the number of variables in the data sets were carried out as specified in Table 13.

Table 13. General data set characteristics

Activity	*Description*	*Result*
Data counting	Unification of the datasets and the count was performed: Facebook Instagram	 132.020 data 27.323 data
Types of values	Categorical, Numeric, Boolean	3
Variables	Followers, Media Type, Time, # hashtags, characters, text/ caption, hashtags, emojis	14

As evidenced in the previous chapter, the need for the project was due to the few studies related to Latin American countries; this chapter sought to address the same need based on the accounts of the thirty-two Colombian governorates for the social networks of Facebook and Instagram. To accomplish it, sixty-four datasets were created, of which most of the extracted variables were numerical values, some categorical variables and others self-calculated to determine the ERP values in each data set (Table 14).

Table 14. Variable description on Facebook and Instagram data sets

Variable	Type of Variable	Description
Account	Categorical	The official account name
Followers	Numerical	Total number of followers in the account
Number of published posts	Numerical	Total number of posts in the account
Post data		
post_url	Categorical	Link that identifies each of the posts
text/ caption	Categorical	Text used per post
date/hour	Numerical	Post creation date - hour

day/week	Numerical	Post creation date - day or week
likes_count	Numerical	Total number of likes per post
Shares	Numerical	Total number of shares per post
comment_count	Numerical	Total number of comments per post
was_live	Boolean	Show if the content was live
emojis_acount	Numerical	Total number of emojis per post
img_count	Numerical	Total number of images per post
hashtag_count	Numerical	Total number of Hashtags used within the text
hashtag_text	Categorical	The text used by hashtag
Media Type	Categorical	Type of content available within the post: image, video, sidecar

Social networks like Facebook and Instagram have stricter privacy policies than other social networks (META, 2022). Therefore, the extraction of the data is under the criteria of responsibility for the use of the data, only public data was used, and there was never personal data collection. For this extraction, the web scraping technique for social networks was used; this allows the automated extraction of information from web pages using programs or scripts. In the case of Facebook and Instagram, by using libraries such as Selenium, BeautifulSoup or Scrapy, navigation through the web pages of these social networks was automated and the desired information was extracted from the HTML of those pages.

No unique coding schemes that required painstaking detail were discovered; it should be clarified that each of the social networks worked on the data sets, and their processing was individual. It is essential to mention that NaN or 0-value values corresponding to empty data for a specific category type were discovered for certain variables. Given how variables are structured in datasets, these are considered present.

3.2 Data processing

Data preparation was one of the most crucial stages in the development of the analytics model, and in this case, it was the stage that took the most time.

The data was prepared at this stage to adapt it to the established data mining techniques. This included choosing the subset of data to use, cleaning it to facilitate interpretation, creating new variables from existing ones, and formatting it as required by the modelling tool and technology based on Facebook and Instagram data.

In cleaning the data, the variables that were closely related to the ERP were prioritised, that is, the variables of interactivity such as likes, shares, and comments and those that are interacting in the incidence of variation of ERP levels such as date/hour, day/week, followers. In addition, it focused on taking categorical variables that allow the text mining process, such as caption/text, comments, hashtags and emojis, to be carried out. The rest of the variables collected were not considered and were discarded from the measurements and the model. At this stage, values such as dates and times were also to homogenised, which were in other time zones by default of social networks, in order to prevent inconsistencies in the modelling since one of the study objectives is to measure the incidences of dates and times at the ERP level for each of the Facebook and Instagram accounts. To evaluate each hypothesis, numerical variables were taken from each data set and the ERP estimate was calculated using formula (1).

Python libraries such as Numpy and Pandas were used for numerical data processing and were widely used in data analysis projects. Numpy offers fast mathematical and statistical operations on multidimensional arrays, while Pandas builds on NumPy and provides high-level data structures called DataFrames and Series. NumPy is an entire library for efficient numerical processing, while Pandas is more specific for tabular data analysis and manipulation. The matplotlib and seaborn libraries were used in data visualisation, which is the most used in data projects. Matplotlib is a more general and flexible plotting library, while Seaborn focuses on statistical visualisations and offers a more simplified and streamlined interface for creating attractive graphs.

On the other hand, for the processing of categorical variables such as caption/text, comments, hashtags and emojis, Natural Language Processing (NPL) techniques were used, which offer functionalities to work with unstructured text and extract meaningful information. Among the functionalities used in the study was: Tokenization, which allowed dividing the text into smaller units, such as words or phrases, which allowed counting words and identifying word patterns. In addition, lemmatisation and stemming were used to reduce words to their base form (lemma) or root, respectively, and it was possible to group similar words and reduce the dimensionality of the vocabulary. These techniques also allowed the processing of hashtags and emojis in processing the text/caption variable. Finally, sentiment analysis was carried out, where the polarity (positive, negative or neutral) of the texts in each data set for Facebook and Instagram was determined.

Similarly, the Python Toolkit for Sentiment Analysis and SocialNLP tasks libraries were used to process categorical variables. This offers pre-trained tools and models to perform sentiment analysis, emotion detection, toxicity, irony and sarcasm, as well as analysis of aspects and opinions in the context of social media, and from which it was possible to look at the incidents in the ERP on use of emojis and hashtags in Facebook and Instagram texts.

3.3 Predictive model for ERP on Facebook and Instagram

The modelling phase was carried out in several iterations. Several models were compiled with the default parameters, and the necessary parametric adjustments were made to meet the previously established business intelligence success criteria. Before the construction of these models, the evaluation methods for the models were determined based on the characteristics of the data and the characteristics of precision and reliability that were to be achieved. As explained in the previous chapter, a random model fitting process, available in the Scikit-learn pipeline module, was used to determine the ideal configuration for each algorithm, fitting it based on the variations of its hyperparameters. This process ensured the best possible results (Lasso et al., 2020).

On the other hand, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (Light GBM), Gradient Boosting Regressor, Random Forest Regressor and Neural Network were the algorithms tested. Mean absolute error (MAE) was used as the evaluation metric, and cross-validation was performed to determine how close the prediction was to the final results for each of the resulting models (Raschka, 2018). Therefore, the ideal parameters for each algorithm used in the data set and its MAE were determined. In addition, the same algorithms were used to evaluate the original (non-reduced) data set and the MAE was obtained. Finally, the algorithm and data subset with the lowest MAE were chosen.

To analyse the importance of each feature in the proposed model, the SHAP library for Python was used. This allowed estimating the SHAP values of each variable, which represent their impact on the final value of a prediction in a particular model from the analysis of the decrease in the performance of a model under different values of a variable (Meng et al., 2020). To illustrate the impact of each variable on the model output, dependency plots and a summary of SHAP values (Lundberg et al., 2020) were used. In addition, understanding how each variable contributed to the construction of the model was achieved, which improved the ERP calculation.

4. Results

Data mining on social networks like Facebook and Instagram using Python libraries is usually done through their respective APIs (Application Programming Interfaces). These APIs allow access to the users' public data and obtain information such as publications, profiles, images, comments, and followers. Data sets were extracted for 32 Facebook and 32 Instagram accounts that belonged to the Colombian governorates in one year, between January 2021 and January 2022. As a result of this extraction exercise, 132,020 records were obtained for Facebook and 27,323 records for Instagram. Likewise, three types of values (categorical, numeric and boolean) and 14 types of variables (Followers, Medi Type, hashtag, caption/text, emojis, ads, shares, likes, time, date, and day) were obtained for each data set.

In the development of the incidences of each variable with the ERP, it was evidenced that the number of Facebook account followers can have a significant impact. It was shown that the ERP of the posts did not increase significantly as the number of followers increased, confirming the H1.

Accounts with more than 150,000 Facebook followers have average ERP values between 0.001 and 0.5 and are considered low in the public sector. In addition, the posting rate for these accounts is between 2,500 and 1,000 per year, as shown in Figure 13.

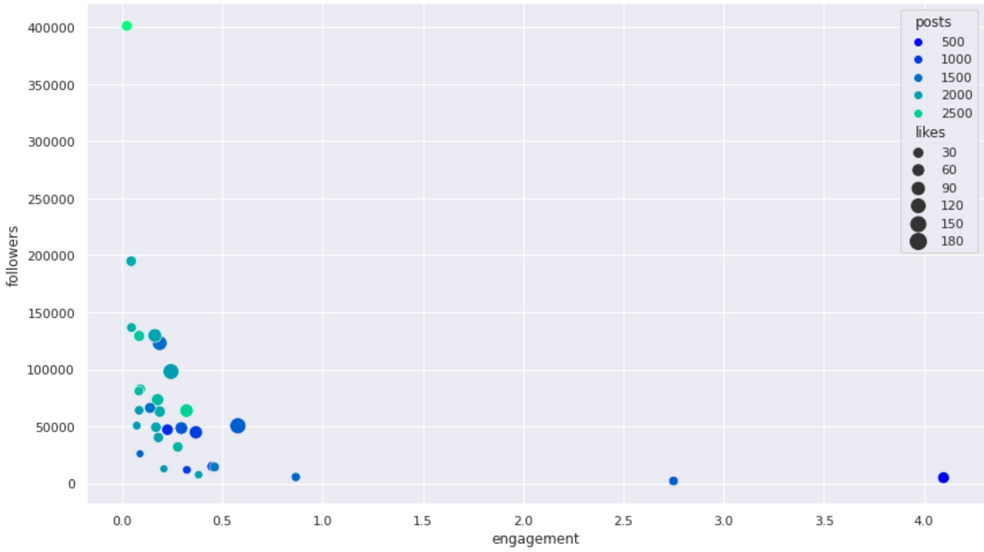


Figure 13. Relationship between Facebook ERP and followers

On the other hand, on Instagram, the behaviour of the data is more dispersed, based on the fact that the number of followers in this social network for the governorates is much lower than networks like Facebook and Twitter. The ERP has better behaviour concerning the posting cadence, higher values are identified when there are between 0 and 1000 posts per account, but there is no evidence of an improvement in the ERP for the number of followers. The atypical data shows that accounts with more than >160,000 followers do not have higher ERP values (Figure 14).

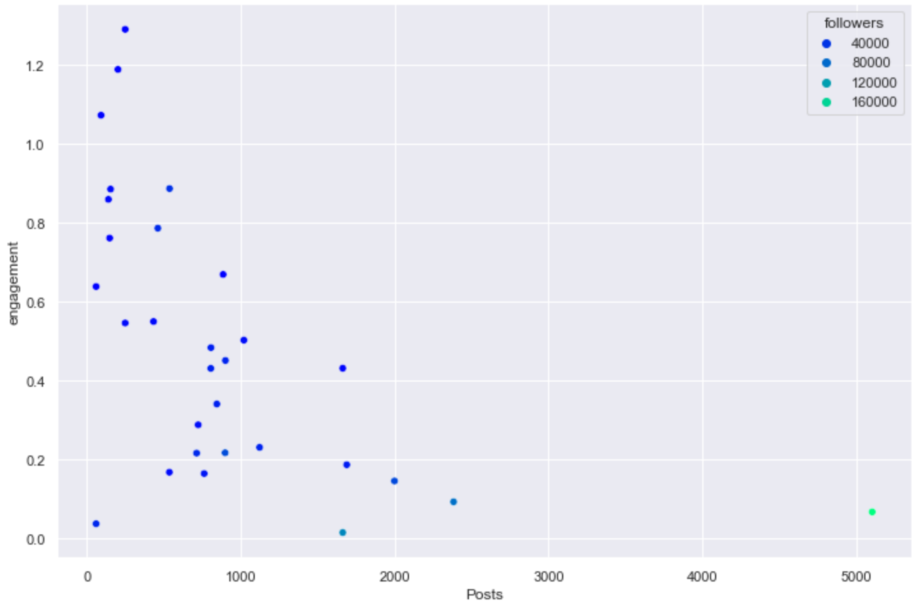


Figure 14. Relationship between Instagram ERP and followers

However, it is shown in Figures 15 and 16 that the type of media (MT) used in the content had a significant impact on the ERP. For the Facebook data set, MT- video is the most used in the content, with an average ERP value of 0.35, followed by MT- image, with an average ERP value of 0.29. On the other hand, the values for Instagram show that the MT that generates the most significant impact on the ERP is the MT- sidecar with a value above 0.4; this content refers to the compendium of multiple images for a single post. In this way, H3 is confirmed, where it is evident that using these multimedia resources improves the ERP on Facebook and Instagram.

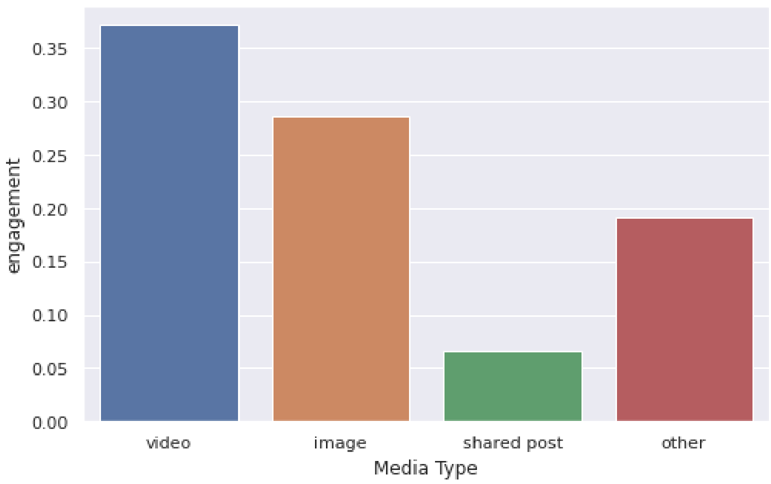


Figure 15. Relationship between Facebook ERP and Media type

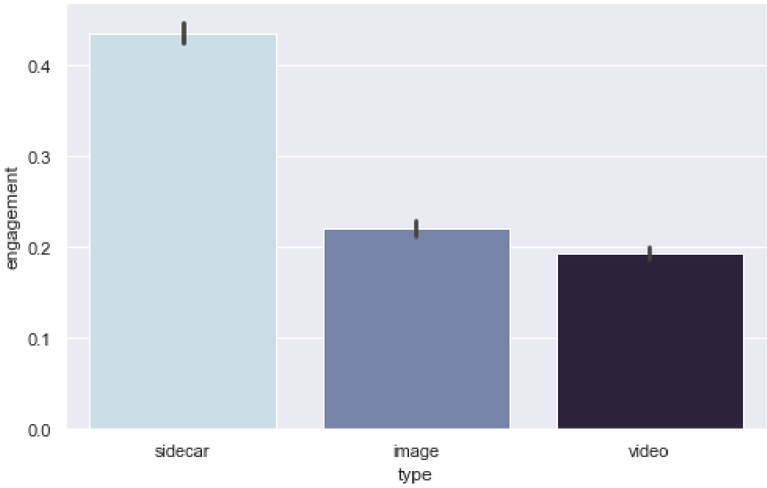


Figure 16. Relationship between Instagram ERP and Media Type

Figure 17 shows the relationship between ERP and posting time (PT) through a heat map to determine the highest ERP points according to posting time per day (PT-*day*) and post time per hour (PT-*hour*). On the map's colour scale, green represented lower values, while red represented higher values. For the Facebook data sets, the measurement range is between 0.4 and > 1.0, while for Instagram, the measurement range is between < 0.025 and > 0.375. Due to the various data extraction techniques, the initial data set contained data from different time zones, which required an adjustment of the PT-day variable to Colombian Standard Time (GMT-5).

The highest value recorded for the ERP in the Facebook data set was 1.057, with a PT-*day* on Monday and TP-*hour* at 17:00, followed by PT-*day* Friday and PT-*hour* at 8:00 with an ERP of 0.086. However, the lowest values registered were PT-*day* Wednesday and PT-*hour* 16:00 of 0.3970. In the Instagram data set, the variables could be related to result in the highest ERP of 0.386 with a TP-*day* Friday and TP-*hour* at 18:00. With the data analysed from Facebook, it is not possible to identify a trend in the variables TP-*day* and TP-*hour*, however, on Instagram it is possible to have a cross-sectional range in TP-*hour* between 5:00 p.m. and 7:00 p.m. and TP-*day* independent (H3).

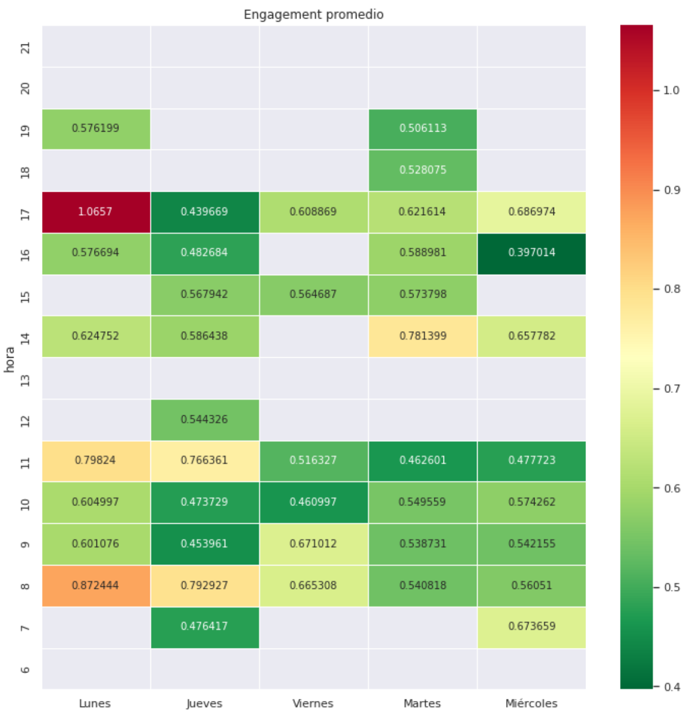


Figure 17. Relationship between TP-*hour*/TP-*day* and the Facebook ERP

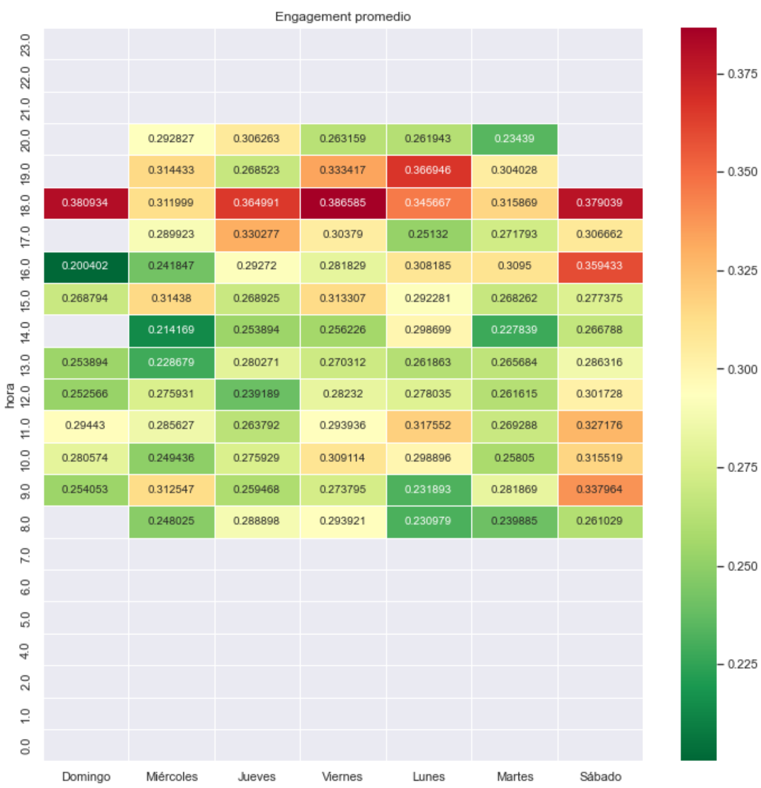


Figure 18. Relationship between TP-*hour*/TP-*day* and the Instagram ERP

The use of hashtags (HT) in the content of Facebook and Instagram has a significant behaviour on the ERP of a user's account. HTs help the audience find content and tag it. The study found that the use of HT has a significant impact on each of the data sets. For Facebook, figure 8 shows that the use of HT in the content between HT 0 and HT 2 values generates average ERP values of 0.5, there is no evidence of a trend, and there is a variable dispersion of data. On the other hand, the data set of Instagram, having HT between 0 and 10, shows average ERP values of 0.6, and it is evident that the ERP improves when it has also HT 20. In this case, H4 applies to the Instagram data set where maximum ERP values are evident concerning the HT, but it is not the same for the Facebook data set where the dispersion of data generates values that may be indifferent to improving the ERP.

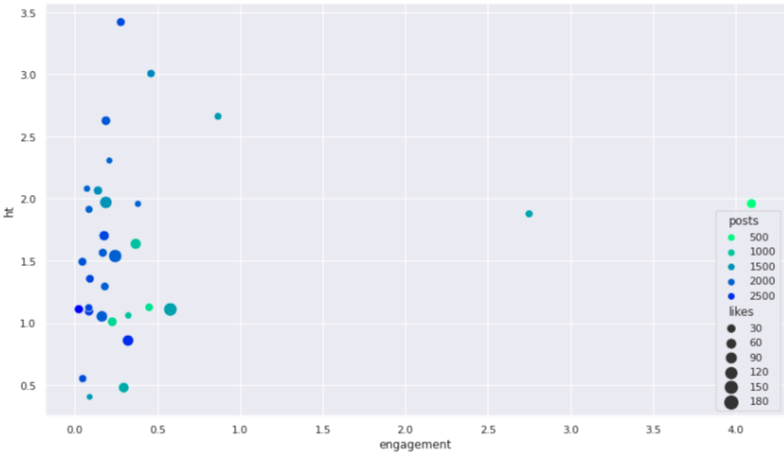


Figure 19. Relationship between hashtags and Facebook ERP

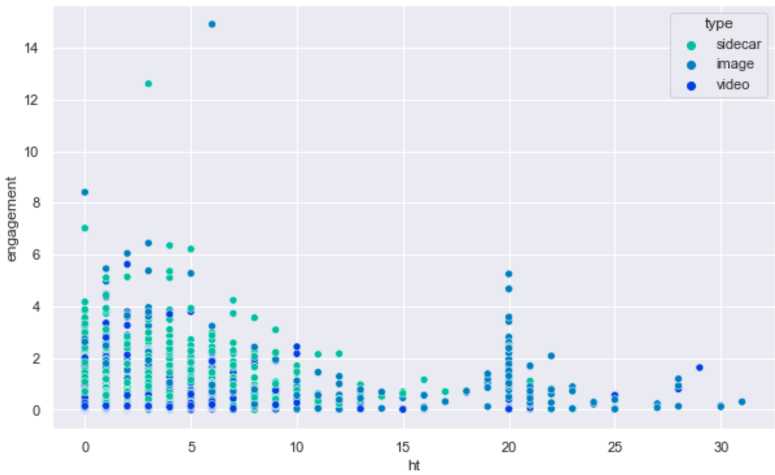


Figure 20. Relationship between hashtags and Instagram ERP
Similarly, using emojis (EM) within Facebook content on Instagram affected ERP behaviour. When its use was from EM 0 to 10, the ERP reached a value

between 0 and 4 (Figure 10) for Facebook. When its use was from EM 0 to 10, the ERP reached a value between 0 and 8 (Figure 11) for Instagram. In addition, it was possible to identify that most of the shared content that uses HT also uses MT-*image* for Facebook and MT-*sidecar* for Instagram.

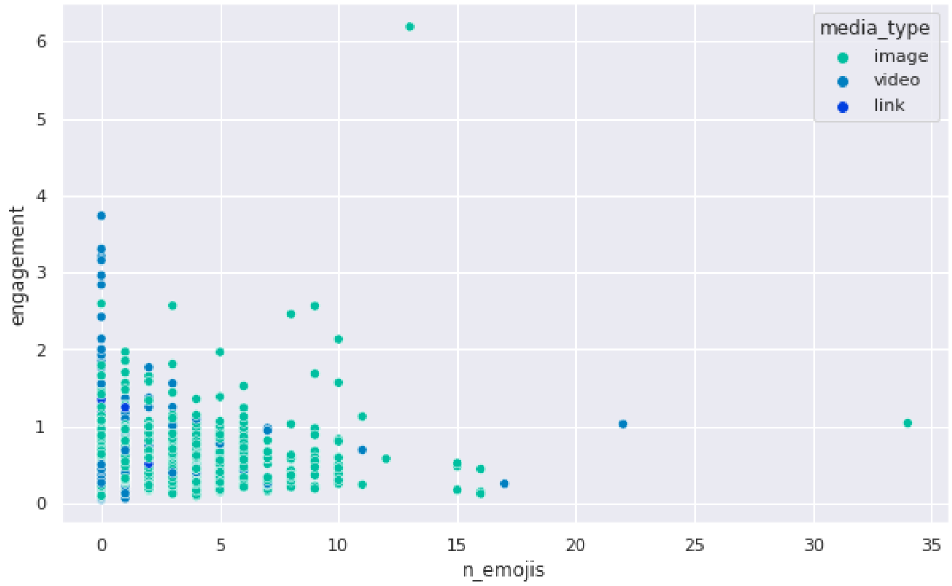


Figure 21. Relationship between emojis and Facebook ERP

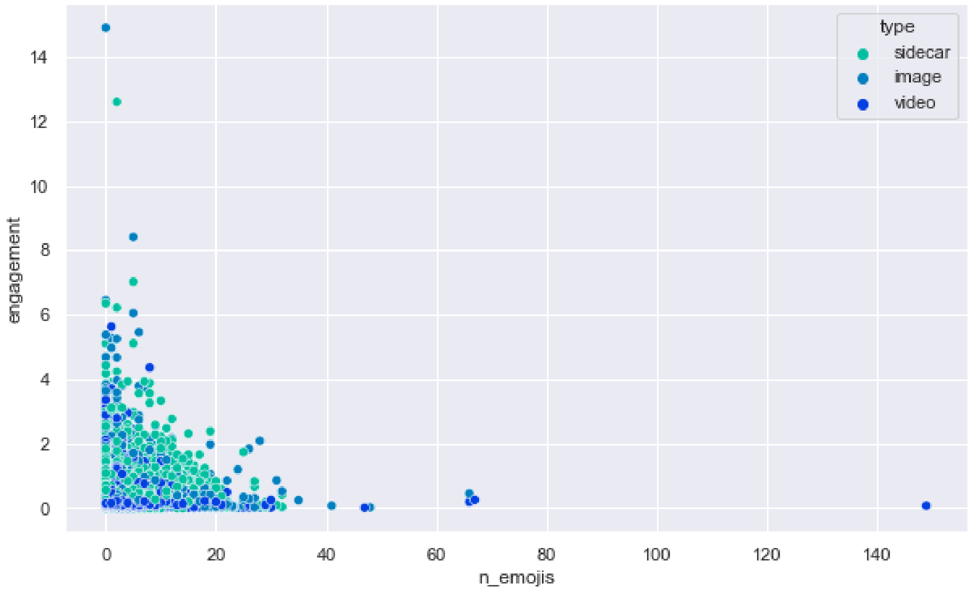


Figure 22. Relationship between emojis and Instagram ERP

Evaluating the H5, sentiment analysis was carried out for the posts with emojis, which underwent transformation and standardisation (Figure 12). As a result, the prevailing sentiment is neutral, with more than >20,000 posts related to information and management texts in each of the governorates. This could be identified when the creation of unigrams ("covid-19", "governance", "tourism"), bigrams ("huilacrece", somosquindio", "vivo") and trigrams ("aestahora", "gobernaciondeputumayo") was made. ", "corazondecolombia"), which mostly showed words related to public policies and the promotion of a government program towards citizens. For this reason, it is considered that the use of elements such as EM or HT does not necessarily generate positive feelings that improve the ERP of the accounts, refuting H5.

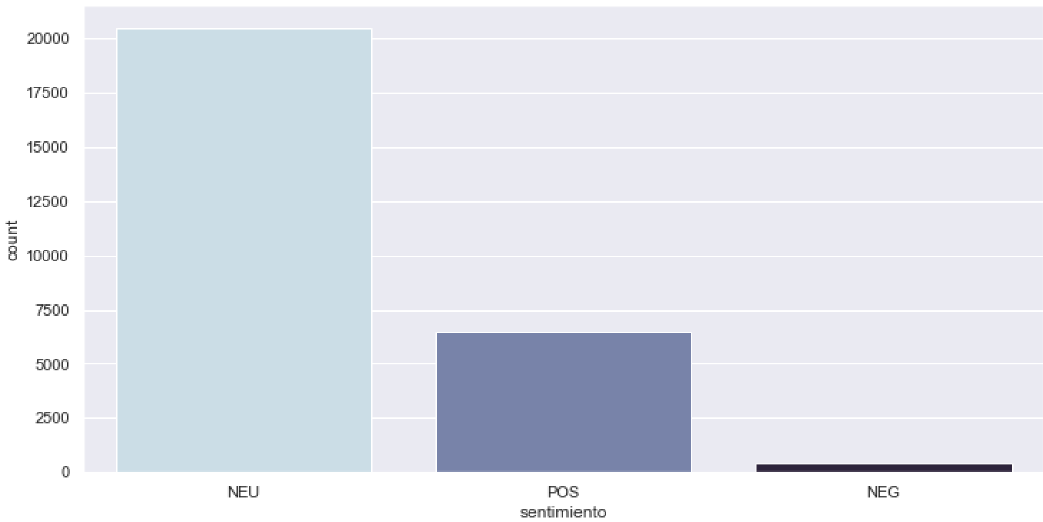


Figure 23. Content sentiment analysis for Instagram

4.1 Predictive model for variables on Facebook and Instagram

For the construction of the model, five regression algorithms were tested to identify the one with the best performance; based on the correlation between the characteristics and the critical values, the three most important were chosen. The algorithm that produced the best results was LightGBM in the two models for both Instagram and Facebook. The results of the variables were a root mean square error (MSE) of 0.1639 and a root mean square error (RMSE) was 0.4049 in the Instagram model. The root mean square error (MSE) of 0.9705 and the root mean square error (RMSE) was 0.9802 in the Facebook model (Table 15).

Table 15. Pipeline results to find the best regression algorithms in Facebook and Instagram.

	Instagram Model	MAE	MSE	RMSE
XG Boost	Extreme Gradient Boosting	0.2389	0.1649	0.4060
LightGBM	Light Gradient Boosting Machine	0.2362	0.1639	0.4049
GBR	Gradient Boosting Regressor	0.2363	0.1664	0.4079
	Facebook Model	**MAE**	**MSE**	**RMSE**
XG Boost	Extreme Gradient Boosting	0.3308	0.9756	0.9877
LightGBM	Light Gradient Boosting Machine	0.3303	0.9705	0.9802
GBR	Gradient Boosting Regressor	0.3313	0.9746	0.9872

The SHAP (Shapley Additive Explanations) values of the characteristics according to the ERP were summarised in the values obtained from the algorithm training (Figure 9). This graph with SHAP values allows to interpret the impact of different characteristics or variables on the result of the machine learning model. Each feature has a colour gradient showing its range of values, with the highest value being red and the lowest being blue. Categorical variables were required to be coded as binary variables for model optimisation and visualisation.

In Figure 24, represented in SHAP values, the violin plot represents the value of each variable on its scale. The impact on the model output is tied to the base value, and its axis shows how the prediction varies above or below this value on a scale from -0.04 to 0.04. When analysing the three main incidence features in the ERP, it is observed that EM (n_emojis) has a high incidence in the model, up to 0.02. In this way, its maximum value (red) is considered as EM (if it has one), and its minimum value (blue) is EM- n (does not have emojis), thus considering that EM-type content increases the expected incidence. Similarly, the TP-hour increases the predicted incidence up to 0.04, allowing us to determine that content published earlier in the day.

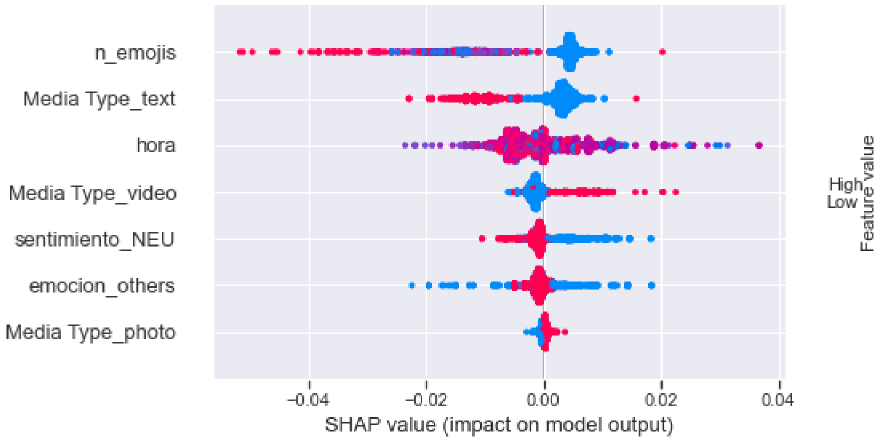


Figure 24. Summary of SHAP values for the Facebook model.

Similarly, in Figure 25, the SHAP values represent the three primary incidence features in the ERP. It is observed that type- sidecar has a high incidence in the model up to 0.04, its maximum value (red) is considered as type -sidecar (if it has one) and its minimum value (blue) does not have this characteristic, thus considering that content with type- sidecar increases the expected incidence. Likewise, the content with the EM variable causes the expected incidence to increase up to 0.02, in addition to having the HT variable where the incidence reaches 1.4.

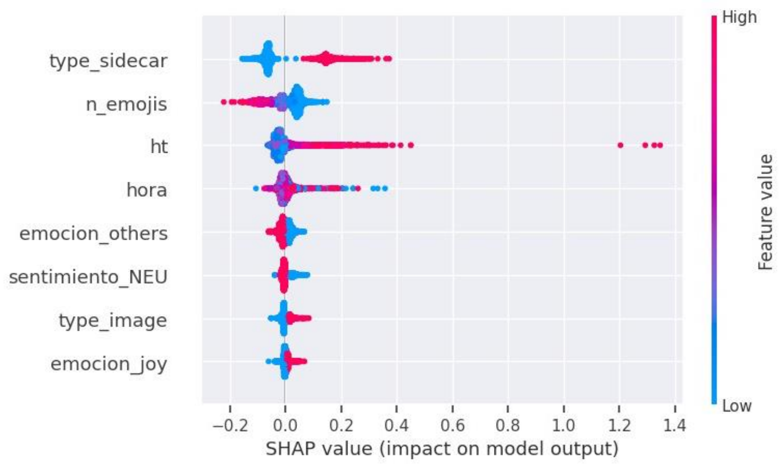


Figure 25. Summary of SHAP values for the Instagram model.

5. Discussion

Local governments have improved communication with their citizens through social networks, allowing a more direct and personalised interaction (Linders, 2020) and generating a sense of transparency and responsibility (Song and Lee, 2015). As a result, social networks have also become a tool to combat misinformation and fake news (Rustad and Sæbø, 2013) and allow the government to produce official digital content. As demonstrated by some of the examined accounts with more structured digital strategies, if the government uses social media effectively, it increases its reach and improves its public image (Choi and Song, 2020). However, it is still evident that governments must strengthen these digital communication strategies to maximise their benefits by creating valuable content and improving ERP levels that facilitate better interactivity with users.

In order to achieve the objective of this study, all the Facebook and Instagram accounts of the Colombian governorates were analysed, representing one hundred per cent of the entire territory. Initial data exploration revealed a significant difference between top departments that produced much content and were highly active on their profiles and those that were further away and had less activity on Facebook and Instagram.

The model made it possible to analyse variables that had not previously been considered in the ERP measurement, which made it possible to evaluate the effectiveness of the MCP. The models previously proposed, such as Bolsón, Royo and Ratkai (2015), do not take variables such as the type of content, the use of media types (image, sidecar, video or text) and the time of posting. For their part, Santoso, Rinjany and Bafadhal (2020) analysed the type of information shared by the Facebook and Twitter accounts, the ones that presented the highest ERP categorised them; in the results, it is possible to show government programs, citizen conditions, new public services or regional activities but does not evaluate factors that influence or affect the ERP, such as the use of emojis or hashtags.

This research showed that Facebook accounts with more than 150,000 followers had shallow ERP values, between 0.001 and 0.5. The results of Instagram showed that it is one of the networks with the fewest followers;

this is due to the low presence that governments have on this platform, and the interactivity with citizens is much more dispersed, which generates ERP values between 0.0 and 1.2. The low values of ERP are transversal to social network platforms, and it is evident that profiles such as governments need to have attention to very well-defined strategies for the CCM. Likewise, the number of posts generated per account does not influence the improvement of the ERP either; accounts with more than 2,500 posts per year did not improve on this measure.

Based on TM variables, many studies (Putrano et al., 2022; Gruzd, Lannigan and Quigley, 2018; Maiano et al., 2021; David, 2010; Haenlein et al., 2020) suggest that this type of content improves the ERP values, and this could be evidenced in the study. For Facebook, MT-video and MT-image showed the highest ERP rates (0.35 and 0.29), unlike content with only text or links. Likewise, the results for Instagram were prioritised using TP- sidecar with values above 0.4, slightly different from Facebook's values. Likewise, the use of hashtags for Facebook turned out to be irrelevant concerning the ERP (0.5) due to the dispersion of the data. Users do not see it as applicable or consider it a search tool on the platform, refuting what it claims Mas-Manchón and Guerrero-Solé (2019). On the other hand, Instagram shows more predominant results and HTs are taken as a tool that allows content to be categorised and tagged. It showed higher ERP results, reaching almost HT 20 with an ERP of 0.6 points.

The posting time (TP) showed that the highest ERP values are TP- day Monday and TP- hour 17:00 for Facebook and TP- day Friday and TP- hour 18:00, being in the hours in which the account presents more excellent activity in the generation of content and interactivity with users. This time differs from what was proposed by Henisa and Wilantika (2021), who state that the days significantly affect it are Wednesday, Saturday and Sunday. However, there is consistency in the publication times that most influence citizen participation are TP- hour 10:00-13:59 (Facebook) and TP- hour 18:00 (Instagram).

Therefore, TP is a variable that does not have consistent Facebook behaviour. However, for Instagram, it is possible to have a trend of ERP improvement at TP- hours 17:00 and 19:00 on weekdays. Likewise, using emojis in content is very common for government accounts. However, its use does not significantly impact the ERP since, through sentiment analysis, it was possible to show that the higher sentiment is NEUTRAL, which allows us to show that it does not influence G2C communication.

6. Conclusion

Since they were created, Facebook and Instagram have been seen as means of disseminating information, so governments began to use them as communication channels for several years, generating G2C or C2G dynamics. The content shared through these platforms supports the actions carried out by governments and determines their presence in social networks and digital channels. The formulas used to evaluate the effectiveness of these contents are transversal to private industries, which limits, to a certain extent, the final objective of the contents. The authors consulted have demonstrated the traditional Engagement Rate per Post (ERP) variables, which include interactivity metrics (likes, shares, and comments) compared to account reach (followers). Hence, this produces unique but limited content effectiveness ratings. A methodology was sought to allow the data project to be organised and adapted to a public content marketing (PCM) process for social networks such as Facebook and Instagram to address this limitation. In addition, creating an analytical model determined which content characteristics improve the effectiveness of the ERP.

The CRISP-DM methodology provided a framework for organising and prioritising data mining tasks. The first stages allowed us to establish the objective and the hypotheses of the study, to understand the gap in the literature on the effectiveness of the content and to determine the technologies for the extraction and evaluation of the quality of the data. Subsequent stages focused on data processing and analysis using machine learning algorithms to identify patterns in the variables that most influenced ERP effectiveness.

The objective of developing an analytical model that would identify which content characteristics improve the effectiveness of the Facebook and Instagram accounts of the Colombian governorates was achieved based on these results and the analysis carried out. The training model was used to analyse variations such as ERP variability, number of followers, type of content (photos, videos, sidecar.), posting hours, hashtags and emojis, and user interactions with the content. Although these variables directly impact the ERP, the content generated by the accounts as their own (Tweet) was the most influential variable to improve the ERP since the ERP decreased by more than 80% when retweets of other accounts, resulting in little interactivity with users. In addition, the use of emojis in the content of these types of accounts was found to have a direct impact on ERPs, which means that government companies need to pay close attention to this.

This study has added to the literature the need to create a complete analytical model to measure the effectiveness of public content marketing (PCM) on Facebook and Instagram, particularly with government entities. Traditional interactivity and reach metrics may need more to accurately and reliably measure MCP effectiveness. ERP can become a small quantity of a fancy metric for channel managers, according to the results obtained, which allowed us to identify the variables that most influence the effectiveness of the content. This study can serve as the basis for future research that seeks to improve the measurement of the effectiveness of public content marketing on Facebook and Instagram, especially to create more complex analytical models that adapt to different contexts and objectives of the accounts.

Chapter 5

Conclusions

Conclusions

Over the past two decades, governments have used information and communications technologies (ICTs) accelerated and significantly to speed up, optimise, adapt, streamline, and make transparent public sector processes or activities. Governments now use digital media more frequently by prioritising communication tactics encouraging meaningful connections with citizens. The top digital media include websites and social networks like Facebook and Twitter that transmit information content, statistical data, and visibility to public projects, online transactions, or open data. This research showed 5.16 billion people or nearly 64.4% of the world's population, were internet users as of January 2022. In Latin America, 453 million internet users or 66.6% of the total population. It is significant to remember that as more individuals acquire access to the internet, these figures are continually changing.

The government's efforts to organise responses to extraordinary occurrences like the pandemic and secure public cooperation during this unprecedented crisis have greatly benefited from using new technologies, mainly social media. These social networks have been essential for keeping societies operating during protracted times of lockdown and fostering cross-sectoral and international collaboration on solutions (Chen et al., 2020). The use of these social networks by governmental organisations offers a unique chance to interact with citizens. Still, it also generates problems and concerns about information openness, tracking usage, and engagement levels.

In the context of public bodies or governments, digital content marketing (DCM) is essential since it offers numerous advantages and opportunities through digital platforms like social media. First of all, it enables these organisations to effectively communicate with the public, providing accurate information about policies, programs, services, and events related to the government. This DCM strategy offers a direct conduit for communication with the target audience and interactivity. Additionally, DCM in social networks has a significant impact on encouraging citizen engagement. Government organisations can encourage citizen collaboration and involvement by encouraging residents to participate in surveys, public forums, and other forms of consultation. However, using these networks for various purposes necessitates measuring effectiveness using various tools and techniques, such as the engagement rate. Digital channels are, by their very nature, a wealth of information that could be better utilised. This study aimed to evaluate how well public entities' content performs on social media platforms like Twitter, Facebook, and Instagram to see which elements need to be improved when using DCM tactics. For this reason, those were the objectives that were developed in each chapter:

- To identify the digital content marketing strategies (DCMS) used on e-Government platforms and create a bibliometric analysis and literature review.
- To analyse the digital content marketing strategies (DCMS) used on Twitter for government entities.
- To analyse the digital content marketing strategies (DCMS) used on Facebook and Instagram for government entities.

The development of the objectives is presented below and concludes with the main results obtained in each chapter.

1. Chapter 2: bibliometric analysis and literature review

The analysis of the co-occurrence of keywords like "Social Media," "e-Government," and "Digital Government" in the majority of the literature consulted, along with others like "Political Communication," "Media," "Engagement," and recurrent use of words like "Twitter" or "Facebook" as primary communication channels for public entities, revealed the results in the first instance. Any government will benefit from the implications of these findings to adequately build its e-government systems and spread them through the relevant social media platforms for the broadest reach and best use by end users.

Contrarily, a gap was found between terms such as "Content Strategy" or "Content Marketing" associated with content creation from digital channels in the digital government. Its frequency was very low or nearly non-existent, which provided an opportunity to investigate new content marketing research directions. In addition, information transparency plays a crucial role in the data discovered due to the prevalence of social media-related problems.

According to research themes, there has been considerable growth in scientific production over the past ten years, with the most significant increase occurring between 2018 and 2020 (Figure 3). The pandemic in 2020 and the scientific community's interest in examining these developments and topics are tied to the rise in the deployment of technologies by governmental entities. Information Science, Business Economics, and Government Law had the most articles per study field (Table 1), showing that journals or sources of contribution also maintained this association while adding another research area, such as Public Health.

As a result, publications like BMC Public Health and the Journal of Medical Internet Research are among the sources that have made the most contributions. These were considered by their Internet, marketing, new technology, digital communication, and public administration research subareas. Additionally, the study found that the leading institutions that contribute the most to scientific production are located in nations like the United Kingdom, the United States, Spain, and China, with universities that are in the top 50 of the World Ranking, demonstrating the interest of these superpowers in researching subjects like DCM and digital government.

2. Chapter 3: Twitter Analytic Model

The effectiveness of the content is assessed using formulas that cross over into private enterprises, which limits the content's ultimate goal to some extent. The measuring variables of the classic Engagement Rate per Post (ERP), which compares the reach of the account (followers) with interaction metrics (likes, shares, retweets, and comments), have been supported by the literature studied. It produces distinctive content effectiveness ratios, however, they may be more adaptable. To overcome this restriction, we looked for a methodology that would enable us to organize the data project and adapt to a Content Marketing (CM) process in order to construct an analytical model identifying the qualities of content that increase efficacy.

It was feasible to accomplish the goal of developing an analytical model that would identify the qualities of the material that increase the efficiency of the Twitter accounts based on these results and the analysis that was acquired. A variety of factors were examined by the training model, including the type of tweet, the number of followers, the type of content (pictures, videos, or gifs), the time since publication, the use of hashtags and emojis, and user interactions with the material (likes, shares, and comments). Although these factors directly affect the ERP, it was discovered.

The content created by the accounts as their own (Tweet) was also the factor that had the greatest influence on the ERP; in contrast, when retweets were made from other accounts, the ERP plummeted by more than 80%, leading to minimal user interaction. Emojis inside the material for this type of account were also discovered to have a direct impact on the ERP, necessitating specific consideration from government organizations. With the use of this discovery, we were able to determine that there are some aspects that are crucial to users' interaction with the material but have not previously been considered when measuring the ERP.

In addition to measuring the success of public content marketing (PCM) on Twitter, this study has made a significant contribution to the body of literature because it found that PCM is highly replicable in a variety of businesses or industries. The findings showed the factors that have the most impact on the material's performance, indicating that more measures than just reach and interaction may be required to effectively and consistently gauge the success of content marketing in this kind of account.

3. Chapter 4: Facebook and Instagram Analytic Model

Through social networks, local governments have increased connection with their constituents, enabling a more direct and individualized interaction and fostering a sense of accountability and transparency. As a result, social media platforms are now used by the government to provide official digital material and to battle fake news and misinformation. If the government uses social media efficiently, it expands its reach and enhances its public image, as shown by some of the investigated accounts with more structured digital strategies. However, it is still clear that governments must improve ERP levels to enable better user interaction in order to develop these digital communication methods and maximize their benefits.

All of the Colombian governorates' Facebook and Instagram profiles, which represent 100% of the country as a whole, were examined in order to meet the study's goal. Initial data analysis showed a clear distinction between the top departments that produced a lot of content and were very active on their pages and those that were farther away and were less active on Facebook and Instagram. The training model was used to analyse differences in things like ERP variability, follower count, content type (pictures, videos, sidecar), posting hours, hashtags, emoticons, and user interactions with the content. The content created by the accounts as their own (Tweet) was the most important variable to enhance the ERP even if these factors also have a direct impact on the ERP. This is because the ERP plummeted by more than 80% when other accounts' tweets were retweeted, which led to little user interaction. Additionally, it has been discovered that the usage of emojis in the content of these kinds of accounts has a direct impact on ERPs, therefore government companies need to pay particular attention to this.

The necessity for a thorough analytical model to assess the success of public content marketing (PCM) on Facebook and Instagram, particularly with governmental organizations, has been raised by this study in the body of literature. To effectively and consistently assess MCP efficacy, traditional engagement and reach measurements may be insufficient. According to the data, ERP might develop into a minor amount of a fancy statistic for channel managers. This information helped us to pinpoint the factors that have the most impact on the success of the content. Future research that aims to enhance the evaluation of the efficiency of public content marketing on Facebook and Instagram, particularly to develop more complicated analytical models that adapt to various contexts and purposes of the accounts, can build on the findings of this study.

References

Reference list

Adjei-Bamfo, P., Maloreh-Nyamekye, T. and Ahenkan, A. (2019). The role of e-government in sustainable public procurement in developing countries: A systematic literature review. *Resources, Conservation and Recycling*, [online] 142, pp.189–203. doi:https://doi.org/10.1016/j.resconrec.2018.12.001.

Al Aufa, B., Sulistiadi, W. and Djawas, F.A. (2020). Measuring Instagram Activity and Engagement Rate of Hospital: A Comparison Before and During COVID-19 Pandemic. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. doi:https://doi.org/10.1109/isriti51436.2020.9315490.

Allcott, H., Gentzkow, M. and Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, [online] 6(2), p.205316801984855. doi:https://doi.org/10.1177/2053168019848554.

Arroyave Cabrera, J.A. (2020). *Colombia*. [online] Media Landscapes. Available at: https://medialandscapes.org/country/colombia/media/social-networks.

Atif, M. and Franzoni, V. (2022). Tell Me More: Automating Emojis Classification for Better Accessibility and Emotional Context Recognition. *Future Internet*, 14(5), p.142. doi:https://doi.org/10.3390/fi14050142.

Ayvaz, S. and Shiha, M.O. (2017). The Effects of Emoji in Sentiment Analysis. *International Journal of Computer and Electrical Engineering*, 9(1), pp.360–369. doi:https://doi.org/10.17706/ijcee.2017.9.1.360-369.

Azevedo, A. and Santos, M. (2008). KDD, semma and CRISP-DM: A parallel overview. In: *IADIS European Conference on Data Mining 2008*. [online] Amsterdam. Available at: https://www.researchgate.net/publication/220969845_KDD_semma_and_CRISP-DM_A_parallel_overview.

Berger, J. and L. Milkman , K. (2012). What Makes Online Content Viral?

Journal of Marketing Research, [online] pp.192–205. doi:https://doi.org/10.1509/jmr.10.0353.

Bhatt, Y., Ghuman, K. and Dhir , A. (2020). Sustainable manufacturing. Bibliometrics and content analysis. *Journal Of Cleaner Production*, p.260. doi:https://doi.org/10.1016/j.jclepro.2020.120988.

Bijmolt, T.H.A., Leeflang, P.S.H., Block, F., Eisenbeiss, M., Hardie, B.G.S., Lemmens, A. and Saffert, P. (2010). Analytics for Customer Engagement. *Journal of Service Research*, 13(3), pp.341–356. doi:https://doi.org/10.1177/1094670510375603.

Bloomstein, M. (2012). Content Strategy at Work. *Elsevier*, p.101.

Bonsón, E., Perea, D. and Bednárová, M. (2019). Twitter as a tool for citizen engagement: An empirical study of the Andalusian municipalities. *Government Information Quarterly*, 36(3), pp.480–489. doi:https://doi.org/10.1016/j.giq.2019.03.001.

Bonsón, E. and Ratkai, M. (2013). A set of metrics to assess stakeholder engagement and social legitimacy on a corporate Facebook page. *Online Information Review*, 37(5), pp.787–803. doi:https://doi.org/10.1108/oir-03-2012-0054.

Bonsón, E., Royo, S. and Ratkai, M. (2015). Citizens' engagement on local governments' Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe. *Government Information Quarterly*, 32(1), pp.52–62. doi:https://doi.org/10.1016/j.giq.2014.11.001.

Bonsón, E., Torres, L., Royo, S. and Flores, F. (2012). Local e-government 2.0: Social media and corporate transparency in municipalities. *Government Information Quarterly*, 29(2), pp.123–132. doi:https://doi.org/10.1016/j.giq.2011.10.001.

Boutet, I., LeBlanc, M., Chamberland, J.A. and Collin, C.A. (2021). Emojis influence emotional communication, social attributions, and information processing. *Computers in Human Behavior*, 119, p.106722. doi:https://doi.org/10.1016/j.chb.2021.106722.

Bozkurt, S., Gligor, D. and Gligor, N. (2021). Investigating the impact of psychological customer engagement on customer engagement behaviors: the moderating role of customer commitment. *Journal of Marketing Analytics*, 10, pp.408–424. doi:https://doi.org/10.1057/s41270-021-00146-3.

Brodie, R.J., Hollebeek, L.D., Jurić, B. and Ilić, A. (2011). Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research. *Journal of Service Research*, 14(3), pp.252–271.

Calderón-Monge, E. and Ramírez-Hurtado, J.M. (2021). Measuring the consumer engagement related to social media: the case of franchising. *Electronic Commerce Research*, 22. doi:https://doi.org/10.1007/s10660-021-09463-2.

Caribbean, E.C. for L.A. and the (2022). *Latin American and the Caribbean Countries are Highly Committed to Pursuing Digital Government Strategies, but Inclusivity and E-participation Remain a Challenge*. [online] www.cepal.org. Available at: https://www.cepal.org/en/news/latin-american-and-caribbean-countries-are-highly-committed-pursuing-digital-government [Accessed 26 Jun. 2023].

Casaló, L., Flavián, C. and Guinalíu, M. (2007). The impact of participation in virtual brand communities on consumer trust and loyalty. *Online Information Review*, 31(6), pp.775–792. doi:https://doi.org/10.1108/14684520710841766.

Cegarra-Navarro, J.-G., Garcia-Perez, A. and Moreno-Cegarra, J.L. (2014). Technology knowledge and governance: Empowering citizen engagement and participation. *Government Information Quarterly*, 31(4), pp.660–668. doi:https://doi.org/10.1016/j.giq.2014.07.001.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C. and Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS inc*.

Chen, Q., Min, C., Zhang, W., Wang, G., Ma, X. and Evans, R. (2020).

Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. *Computers in Human Behavior*, [online] 110, p.106380. doi:https://doi.org/10.1016/j.chb.2020.106380.

Cheng, J., Danescu-Niculescu-Mizil, C. and Leskovec, J. (2014). How Community Feedback Shapes User Behavior. In: *In Proceedings of the 22nd international conference on World Wide Web*. pp.225–236.

Choi, J.-C. and Song, C. (2020). Factors explaining why some citizens engage in E-participation, while others do not. *Government Information Quarterly*, 37(4), p.101524. doi:https://doi.org/10.1016/j.giq.2020.101524.

Chun, S.A., Shulman, S., Sandoval, R. and Hovy, E. (2010). Government 2.0: Making connections between citizens, data and government. *Information Polity*, 15(1,2), pp.1–9. doi:https://doi.org/10.3233/ip-2010-0205.

CIGI & IPSOS (2019). Internet security & trust: social media, fake news & algorithms. *CIGI-IPSOS global survey internet*, [online] p.19. Available at: https://bit.ly/3pQq28S.

Contreras Orozco, L. (2017). Las redes sociales digitales como factor de proximidad entre la sociedad y los gobiernos subnacionales en México. *RICSH Revista Iberoamericana de las Ciencias Sociales y Humanísticas*, 6(12), pp.225–249. doi:https://doi.org/10.23913/ricsh.v6i12.131.

Criado, J., Guevara-Gómez, A. and Villodre, J. (2020). Using Collaborative Technologies and Social Media to Engage Citizens and Governments during the COVID-19 Crisis. The Case of Spain. *Digital Government: Research and Practice*, 1(4), pp.1–7.

Criado, J.I., Sandoval-Almazan, R. and Gil-Garcia, J.R. (2013). Government innovation through social media. *Government Information Quarterly*, 30(4), pp.319–326. doi:https://doi.org/10.1016/j.giq.2013.10.003.

Cuevas-Molano, E., Matosas-López, L. and Bernal-Bravo, C. (2021). Factors Increasing Consumer Engagement of Branded Content in Instagram. *IEEE Access*, [online] 9, pp.143531–143548.

doi:https://doi.org/10.1109/ACCESS.2021.3121186.

Dagnino, G., Levanti, G., Minà, A. and Picone, P. (2015). Interorganizational network and innovation: a bibliometric study and proposed research agenda. *Journal Of Business & Industrial Marketing*, 30(3/4), pp.354–377. doi:https://doi.org/10.1108/jbim-02-2013-0032.

Dakouan, C. and Benabdelouahed, R. (2019). Inbound Marketing vs. Outbound Marketing: Independent or Complementary Strategies. *Expert Journal of Marketing*, 7(1), pp.1–6.

Datareportal (2022). *DIGITAL 2022: GLOBAL OVERVIEW REPORT*. [online] *Datareportal*. Available at: https://datareportal.com/reports/digital-2022-global-overview-report.

Datareportal (2023). *DIGITAL 2023: GLOBAL OVERVIEW REPORT*. [online] *datareportal.com*. Available at: https://datareportal.com/reports/digital-2023-global-overview-report.

David, G. (2010). Camera phone images, videos and live streaming: a contemporary visual trend. *Visual Studies*, 25(1), pp.89–98. doi:https://doi.org/10.1080/14725861003607017.

De Blasio, E. and Selva, D. (2018). Implementing open government: a qualitative comparative analysis of digital platforms in France, Italy and United Kingdom. *Quality & Quantity*, 53(2), pp.871–896. doi:https://doi.org/10.1007/s11135-018-0793-7.

De Blasio, E. and Selva, D. (2019). Implementing open government: a qualitative comparative analysis of digital platforms in France, Italy, and United Kingdom. *Quality and Quantity*, 53(2), pp.871–896. doi:https://doi.org/10.1007/s11135-018-0793-7.

De Blasio, E. and Sorice, M. (2016). Open Government: A Tool for Democracy? *Medijske Studije*, 7(14), pp.14–31. doi:https://doi.org/10.20901/ms.7.14.3.

de Jong, M.D.T., Neulen, S. and Jansma, S.R. (2019). Citizens' intentions to participate in governmental co-creation initiatives: Comparing three co-creation configurations. *Government Information Quarterly*, 36(3), pp.490–500. doi:https://doi.org/10.1016/j.giq.2019.04.003.

Deighton, J. and Kornfeld, L. (2009). Interactivity's unanticipated consequences for marketers and marketing. *Journal of Interactive Marketing*, 23(no. 1), pp.4–10.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, pp.285–296.

Dwivedi, Y., Rana, N., Janssen , M., Lal, B., Williams, M. and Clement, M. (2017). An empirical validation of a unified model of electronic government adoption (UMEGA). *Government Information Quarterly*, 34(2), pp.211–230. doi:https://doi.org/10.1016/j.giq.2017.03.001.

Edmonds, W. and Kennedy, T. (2017). An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods. *SAGE Publications*, *Inc*, 2nd ed. doi:http://www.doi.org/10.4135/9781071802779.

Ekman, J. and Amnå, E. (2012). Political participation and civic engagement: Towards a new typology. *Human Affairs*, 22(3), pp.283–300. doi:https://doi.org/10.2478/s13374-012-0024-1.

Ellison, N. and Hardey, M. (2013). Social Media and Local Government: Citizenship, Consumption and Democracy. *Local Government Studies*, 40(1), pp.21–40. doi:https://doi.org/10.1080/03003930.2013.799066.

Eltantawy, N. and Wiest, J. (2011). Social Media in the Egyptian Revolution: Reconsidering Resource Mobilization Theory. *International Journal of Communication*, 5(1), pp.1207–1224. doi:https://doi.org/1932%E2%80%938036/2011FEA1207.

Enria, L., Waterlow, N., Rogers, N.T., Brindle, H., Lal, S., Eggo, R.M., Lees, S. and Roberts, C. h (2020). Trust and Transparency in times of Crisis: Results from an Online Survey During the First Wave (April 2020) of the COVID-19 Epidemic in the UK. *PLoS ONE*. [online] doi:https://doi.org/10.1101/2020.09.01.20183822.

Eom, S.-J., Hwang, H. and Kim, J.H. (2018). Can social media increase government responsiveness? A case study of Seoul, Korea. *Government Information Quarterly*, 35(1), pp.109–122. doi:https://doi.org/10.1016/j.giq.2017.10.002.

Erle, T.M., Schmid, K., Goslar, S.H. and Martin, J.D. (2021). Emojis as social information in digital communication. *Emotion*. doi:https://doi.org/10.1037/emo0000992.

Eslami, S.P., Ghasemaghaei, M. and Hassanein, K. (2021). Understanding consumer engagement in social media: The role of product lifecycle. *Decision Support Systems*, 162, p.113707. doi:https://doi.org/10.1016/j.dss.2021.113707.

Falco, E. and Kleinhans, R. (2018). Beyond technology: Identifying local government challenges for using digital platforms for citizen engagement. *International Journal of Information Management*, 40, pp.17–20. doi:https://doi.org/10.1016/j.ijinfomgt.2018.01.007.

Fernández Cano, A. and Bueno Sánchez, A. (1998). Síntesis de estudios bibliométricos españoles en educación. Una dimensión evaluativa. *Revista Española De Documentación Científica*, 21(3), pp.269–285. doi:https://doi.org/10.3989/redc.1998.v21.i3.356.

Feroz Khan, G., Young Yoon, H., Kim, J. and Woo Park, H. (2014). From e-government to social government: Twitter use by Korea’s central government. *Online Information Review*, 38(1), pp.95–113. doi:https://doi.org/10.1108/oir-09-2012-0162.

Finquelievich, S. and Prince, A. (2013). El Gobierno Electrónico como factor de desarrollo urbano-regional. *Avatares de la comunicación y la cultura*, 2(1853-5925).

Flynn, D.J., Nyhan, B. and Reifler, J. (2017). The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs about Politics. *Political Psychology*, [online] 38(S1), pp.127–150. doi:https://doi.org/10.1111/pops.12394.

Fugate, J.M.B. and Franco, C.L. (2021). Implications for Emotion: Using

Anatomically Based Facial Coding to Compare Emoji Faces Across Platforms. *Frontiers in Psychology*, 12. doi:https://doi.org/10.3389/fpsyg.2021.605928.

Galvez Torres, E., Cruz Alfaro, M., Cespedes Blanco, C., Raymundo, C., Mamani-Macedo, N. and Dominguez, F. (2020). B2B Marketing Method Adapted to Sales Improvement Through the Implementation of ABC Classification Tool and Inbound Marketing in SMEs. *Springer*, 1209. doi:https://doi.org/10.1007/978-3-030-50791-6_21.

Gálvez, C. (2018). Análisis de co-palabras aplicado a los artículos muy citados en Biblioteconomía y Ciencias de la Información (2007-2017). *Transinformação*, 30(3), pp.277–286. doi:http://dx.doi.org/10.1590/2318-0889201800030000.

Gao, X. and Lee, J. (2017). E-government services and social media adoption: Experience of small local governments in Nebraska state. *Government Information Quarterly*, 34(4), pp.627–634. doi:https://doi.org/10.1016/j.giq.2017.09.005.

Gil de Zúñiga, H., Jung, N. and Valenzuela, S. (2012). Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication*, [online] 17(3), pp.319–336. doi:https://doi.org/10.1111/j.1083-6101.2012.01574.x.

Godin, S. (2007). *Permission Marketing*. 1st ed ed. London: Pocket, p.34.

Gordon, E., Baldwin-Philippi, J. and Balestra, M. (2013). Why We Engage: How Theories of Human Behavior Contribute to Our Understanding of Civic Engagement in a Digital Era. *SSRN Electronic Journal*, 1(1). doi:https://doi.org/10.2139/ssrn.2343762.

Gruzd, A., Lannigan, J. and Quigley, K. (2018). Examining government cross-platform engagement in social media: Instagram vs Twitter and the big lift project. *Government Information Quarterly*, 35(4), pp.579–587. doi:https://doi.org/10.1016/j.giq.2018.09.005.

Guillamón, M.-D., Ríos, A.-M., Gesuele, B. and Metallo, C. (2016). Factors influencing social media use in local governments: The case of Italy and Spain. *Government Information Quarterly*, 33(3), pp.460–471. doi:https://doi.org/10.1016/j.giq.2016.06.005.

Haenlein, M., Anadol, E., Farnsworth, T., Hugo, H., Hunichen, J. and Welte, D. (2020). Navigating the New Era of Influencer Marketing: How to Be Successful on Instagram, TikTok, & Co. *California Management Review*, [online] 63(1), pp.5–25. doi:https://doi.org/10.1177/0008125620958166.

Halvorson, K., Rach, M. and Cancilla, S. (2012). *Content strategy for the web*. *New Riders*, New Riders, p.13.

Handley, A. and Chapman, C. (2011). *Content rules*. John Wiley & Sons, p.21.

Hantrais, L., Allin, P., Kritikos, M., Sogomonjan, M., Anand, P., Livingstone, S., Williams, M. and Innes, M. (2020). Covid-19 and the digital revolution. *Contemporary Social Science*, 16(2), pp.256–270. doi:https://doi.org/10.1080/21582041.2020.1833234.

Haro-de-Rosario, A., Sáez-Martín, A. and del Carmen Caba-Pérez, M. (2016). Using social media to enhance citizen engagement with local government: Twitter or Facebook? *New Media & Society*, [online] 20(1), pp.29–49. doi:https://doi.org/10.1177/1461444816645652.

Hemphill, L., Culotta, A. and Heston, M. (2013). Framing in Social Media: How the US Congress Uses Twitter Hashtags to Frame Political Issues. *SSRN Electronic Journal*. [online] doi:https://doi.org/10.2139/ssrn.2317335.

Henisa , D. and Wilantika, N. (2022). Content Characteristics of Government Social Media and The Impact on Citizen Engagement Rate. In: *IEEE Xplore*. International conference on Informatics, Multimedia, Cyber and Information System.

Hollebeek, L.D. and Macky, K. (2019). Digital Content Marketing's Role in Fostering Consumer Engagement, Trust, and Value: Framework, Fundamental Propositions, and Implications. *Journal of Interactive Marketing*, 45(1), pp.27–41.

Holliman, G. and Rowley, J. (2014). Business to business digital content marketing: marketers' perceptions of best practice. *Journal of Research in Interactive Marketing*, 8(4), pp.269–293. doi:https://doi.org/10.1108/jrim-02-2014-0013.

Hutchins, J. and Rodriguez, D. (2018). The soft side of branding: Leveraging emotional intelligence. *Journal of Business & Industrial Marketing*, 1, pp.17–25. doi:https://doi.org/,%20https://doi.org/10.1108/JBIM-02-2017-0053.

Jahn, B. and Kunz, W.H. (2012). How to Transform Consumers into Fans of Your Brand. *SSRN Electronic Journal*, 23(3). doi:https://doi.org/10.2139/ssrn.2307093.

Joo, S., Lu, K. and Lee, T. (2020). Analysis of content topics, user engagement and library factors in public library social media based on text mining. *Online Information Review*, 44(1), pp.258–277. doi:https://doi.org/10.1108/oir-11-2018-0345.

José, J., Juan Pedro Giudici and Luque, F.M. (2021). *pysentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP tasks.* [online] Research Gate. Available at: https://www.researchgate.net/publication/352505261_pysentimiento_A_Python_Toolkit_for_Sentiment_Analysis_and_SocialNLP_tasks [Accessed 4 Jun. 2023].

Jutkowitz, A. (2014). The Content Marketing Revolution. *Harvard Business Review*. [online] Available at: https://bit.ly/3aAAjS5 [Accessed 1 Oct. 2021].

Kapetanaki, A.B., Bertele, K. and Brennan, R. (2017). Social Marketing Through Social Media: The Effects of User-Generated Content. *SSRN Electronic Journal* . doi:https://doi.org/10.2139/ssrn.3112847.

Kee, A. and Yazdanifard, R. (2015). The Review of Content Marketing as a New Trend in Marketing Practices. *International Journal of Management, Accounting and Economics*, 2(9), pp.1055–1064.

Kemp, S. (2022). *Digital 2022: Colombia*. [online] DataReportal – Global

Digital Insights. Available at: https://datareportal.com/reports/digital-2022-colombia.

Khalid, S., Khalil, T. and Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. Science and Information Conference. pp.372–378. doi:https://doi.org/10.1109/SAI.2014.6918213..

Khan, G.F. (2017). *Social Media for Government Implementing and Managing a Participatory, Open, and Collaborative Government*. Singapore Springer Ann Arbor, Michigan Proquest.

Koob, C. (2021). Determinants of content marketing effectiveness: Conceptual framework and empirical findings from a managerial perspective. *PLOS ONE*, [online] 16(4), p.e0249457. doi:https://doi.org/10.1371/journal.pone.0249457.

Kotler, P., Kartajaya, H. and Setiawan, I. (2017). *Marketing 4.0*. 1st ed. MT Biznes, p.120.

Kumar, P., Sharma, A. and Salo, J. (2019). A bibliometric analysis of extended key account management literature. *Industrial Marketing Management*, 82, pp.276–292. doi:https://doi.org/10.1016/j.indmarman.2019.01.006.

Kunal, S., Saha, A., Varma, A. and Tiwari, V. (2018). Textual Dissection of Live Twitter Reviews using Naive Bayes. *Procedia Computer Science*, [online] 132, pp.307–313. doi:https://doi.org/10.1016/j.procs.2018.05.182.

Lappas, G., Triantafillidou, A., Deligiaouri, A. and Kleftodimos, A. (2018). Facebook Content Strategies and Citizens' Online Engagement: The Case of Greek Local Governments. *The Review of Socionetwork Strategies*, 12(1), pp.1–20. doi:https://doi.org/10.1007/s12626-018-0017-6.

Lasso, E., Corrales, D.C., Avelino, J., de Melo Virginio Filho, E. and Corrales, J.C. (2020). Discovering weather periods and crop properties favorable for coffee rust incidence from feature selection approaches. *Computers and Electronics in Agriculture*, 176(1), p.105640.

doi:https://doi.org/10.1016/j.compag.2020.105640.

Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J. and Zittrain, J.L. (2018). The science of fake news. *Science*, 359(6380), pp.1094–1096. doi:https://doi.org/10.1126/science.aao2998.

LeCompte, T. and Chen, J. (2017). Sentiment Analysis of Tweets Including Emoji Data. *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*. doi:https://doi.org/10.1109/csci.2017.137.

Lei, S.S.I., Pratt, S. and Wang, D. (2016). Factors influencing customer engagement with branded content in the social network sites of integrated resorts. *Asia Pacific Journal of Tourism Research*, 22(3), pp.316–328. doi:https://doi.org/10.1080/10941665.2016.1250792.

Lev-On, A. and Steinfeld, N. (2015). Local engagement online: Municipal Facebook pages as hubs of interaction. *Government Information Quarterly*, 32(3), pp.299–307. doi:https://doi.org/10.1016/j.giq.2015.05.007.

Limniou, M., Ascroft, Y. and McLean, S. (2021). Differences between Facebook and Instagram Usage in Regard to Problematic Use and Well-Being. *Journal of Technology in Behavioral Science*. doi:https://doi.org/10.1007/s41347-021-00229-z.

Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly*, [online] 29(4), pp.446–454. doi:https://doi.org/10.1016/j.giq.2012.06.003.

Líppez De Castro, S. and García Alonso, R. (2016). Ciudadanos y gobierno electrónico: la orientación al ciudadano de los sitios Web municipales en Colombia para la promoción de la participación. *Universitas Humanística*, 82(82). doi:https://doi.org/10.11144/javeriana.uh82.cgeo.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), pp.56–67. doi:https://doi.org/10.1038/s42256-019-0138-9.

Lundberg, S.M. and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. [online] 31st Conference on Neural Information Processing Systems . Available at: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf [Accessed 22 Oct. 2022].

MacKay, M., Colangeli, T., Gosselin, S., Neumann, S. and Papadopoulos, A. (2022). Engagement Analysis of Canadian Public Health and News Media Facebook Posts and Sentiment Analysis of Corresponding Comments during COVID-19. *Psych*, 4(1), pp.60–70. doi:https://doi.org/10.3390/psych4010005.

Maiano, L., Amerini, I., Ricciardi Celsi, L. and Anagnostopoulos, A. (2021). Identification of Social-Media Platform of Videos through the Use of Shared Features. *Journal of Imaging*, 7(8), p.140. doi:https://doi.org/10.3390/jimaging7080140.

Martínez, F., Merigó, J., Valenzuela, L. and Nicolás, C. (2018). Fifty years of the European Journal of Marketing: a bibliometric analysis. *European Journal of Marketing*, 52(1/2), pp.439–468. doi:https://doi.org/10.1108/ejm-11-2017-0853.

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez Orallo, J., Kull, M., Lachiche, N., Ramirez Quintana, M.J. and Flach, P.A. (2020). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, [online] 33(8), pp.1–1. doi:https://doi.org/10.1109/tkde.2019.2962680.

Mas-Manchón, L. and Guerrero-Solé, F. (2019). The use of hashtags as a political branding strategy / El uso de los hashtags como una estrategia de marca política. *Revista Internacional de Relaciones Públicas,* 9(17). doi:https://doi.org/10.5783/rirp-17-2019-02-05-24.

Matsumoto, K., Yoshida, M. and Kita, K. (2018). Classification of Emoji Categories from Tweet Based on Deep Neural Networks. In: *I Association for Computing Machinery*. NLPIR 2018: 2nd International Conference on Natural Language Processing and Information Retrieval . United States: Association for Computing MachineryNew York, United States. doi:https://doi.org/10.1145/3278293.3278306.

Mayor, E. and Bietti, L.M. (2021). Twitter, time and emotions. *Royal Society Open Science*, 8(5), p.201900. doi:https://doi.org/10.1098/rsos.201900.

Mencia, I. (2020). *Social Media in Colombia: The Full Overview to Begin Marketing Plans*. [online] Colibri Content. Available at: https://www.colibricontent.com/social-media-in-colombia/.

Meng, Y., Yang, N., Qian, Z. and Zhang, G. (2020). What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(3), pp.466–490. doi:https://doi.org/10.3390/jtaer16030029.

Merigó, J., Mas-Tur, A., Roig-Tierno, N. and Ribeiro-Soriano, D. (2015). bibliometric overview of the Journal of Business Research between 1973 and 2014. *Journal of Business Research*, 68(12), pp.2645–2653. doi:https://doi.org/10.1016/j.jbusres.2015.04.006.

Merigó, J. and Yang, J. (2017). A bibliometric analysis of operations research and management science. *Omega*, 73, pp.37–48. doi:https://doi.org/10.1016/j.omega.2016.12.004.

META (2022). *Política de privacidad de Meta: cómo Meta recopila y usa los datos de usuarios*. [online] es-la.facebook.com. Available at: https://es-la.facebook.com/privacy/policy/ [Accessed 27 Jun. 2022].

MINTIC (2017). *El Plan Vive Digital 2014-2018 - MINTIC - Vive Digital*. [online] mintic.gov.co. Available at: https://mintic.gov.co/portal/vivedigital/612/w3-article-19654.html [Accessed 2 Mar. 2023].

MINTIC (2021). *Hablemos de redes sociales en entidades del Estado*. [online] Gobierno en redes. Available at:

https://www.gobiernoenredes.gov.co/hablemos-redes-sociales-entidades-del-estado/#:~:text=%E2%9C%85%20Para%20los%20colombianos%2C%20las [Accessed 19 Jul. 2022].

Mulgan, R. (2014). *Making Open Government Work*. 1st ed. Bloomsbury Publishing.

Mulyono, B., Affandi, I., Suryadi, K. and Darmawan, C. (2022). Online civic engagement: Fostering citizen engagement through social media. *Jurnal Civics: Media Kajian Kewarganegaraan*, 19(1), pp.75–85. doi:https://doi.org/10.21831/jc.v19i1.49723.

Muñoz-Expósito, M., Oviedo-García, M.Á. and Castellanos-Verdugo, M. (2017). How to measure engagement in Twitter: advancing a metric. *Internet Research*, 27(5), pp.1122–1148. doi:https://doi.org/10.1108/intr-06-2016-0170.

Naser, A. and Concha, G. (2011). *El gobierno electronico en la gestion publica*. [online] *SERIES DE LA CEPAL GESTIÓN PÚBLICA*. CEPAL. Available at: https://www.cepal.org/es/publicaciones/7330-gobierno-electronico-la-gestion-publica.

Neel, L.A.G., McKechnie, J.G., Robus, C.M. and Hand, C.J. (2023). Emoji Alter the Perception of Emotion in Affectively Neutral Text messages. *Journal of Nonverbal Behavior*. doi:https://doi.org/10.1007/s10919-022-00421-6.

Onwuegbuzie, A. and Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In: *Handbook of mixed methods in social and behavioural research*. Sage, pp.351–384.

Padgett, C., Moffitt, R.L. and Grieve, R. (2021). More than words: Using digital cues to enhance student perceptions of online assignment feedback. *The Internet and Higher Education*, 49, p.100789. doi:https://doi.org/10.1016/j.iheduc.2020.100789.

Paesbrugghe, B., Sharma, A., Rangarajan, D. and Syam, N. (2018). Personal selling and the purchasing function: where do we go from here? *Journal*

Of Personal Selling & Sales Management, 38(1), pp.123–143. doi:https://doi.org/10.1080/08853134.2018.1425881.

Palacios, H.J.G., Pantoja, G.A.H., Navarro, A.A.M., Puetaman, I.M.A. and Toledo, R.A.J. (2016). Comparative between CRISP-DM and SEMMA for data cleaning of MODIS products in a study of land use and land cover change. *2016 IEEE 11th Colombian Computing Conference, CCC 2016 - Conference Proceedings*. doi:https://doi.org/10.1109/columbiancc.2016.7750789.

Pang, P.C.-I., Cai, Q., Jiang, W. and Chan, K.S. (2021). Engagement of Government Social Media on Facebook during the COVID-19 Pandemic in Macao. *International Journal of Environmental Research and Public Health*, 18(7), p.3508. doi:https://doi.org/10.3390/ijerph18073508.

Patrutiu, L. (2016). Inbound Marketing - the most important digital marketing strategy. *Bulletin of the Transilvania University of Braşov*, 9(2), pp.61–68.

Peker, S. and Kart , Ö. (2023). Transactional data-based customer segmentation applying CRISP-DM methodology: A systematic review. *Journal of Data, Information and Management*, 5, pp.1–21. doi:https://doi.org/10.1007/s42488-023-00085-x.

Plano-Clark, V. and Ivankova, N. (2016). What is mixed methods research?: considering how mixed methods research is defined. In: *Mixed methods research: A guide to the field*. Sage, pp.55–78.

Pulizzi, J. (2012a). *Content Marketing Definition - Examples*. [online] Content Marketing Institute. Available at: http://contentmarketinginstitute.com/2012/06/content-marketing-definition/ [Accessed May 2022].

Pulizzi, J. (2012b). The Rise of Storytelling as the New Marketing. *Publishing Research Quarterly*, [online] 28(2), pp.116–123. doi:https://doi.org/10.1007/s12109-012-9264-5.

Putranto, H.A., Rizaldi, T., Riskiawan, H.Y., Setyohadi, D.P.S., Atmadji, E.S.J. and Nuryanto, I.H. (2022). Measurement of Engagement Rate on

Instagram for Business Marketing . In: *Proceedings - IEIT 2022*. : 2022 International Conference on Electrical and Information Technology. pp.317–321.

Ramirez-Madrid, J.P., Escobar-Sierra, M., Lans-Vargas, I. and Montes Hincapie, J.M. (2022). Government influence on e-government adoption by citizens in Colombia: Empirical evidence in a Latin American context. *PLOS ONE*, 17(2), p.e0264495. doi:https://doi.org/10.1371/journal.pone.0264495.

Ramírez-Montoya, M. and Lugo-Ocando, J. (2020). Systematic review of mixed methods in the framework of educational innovation. *Comunicar*, 28(65), pp.9–20. doi:https://doi.org/10.3916/c65-2020-01.

Randhawa, K., Wilden, R. and Hohberger, J. (2016). A Bibliometric Review of Open Innovation: Setting a Research Agenda. *Journal Of Product Innovation Management*, 33(6), pp.750–772. doi:https://doi.org/10.1111/jpim.12312.

Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, 3(24), p.638. doi:https://doi.org/10.21105/joss.00638.

Rodriguez Leon , C. and Garcia Lorenzo, M.M. (2016). Adaptation To A Methodology Of Data Mining for Applying to Un-Su-Pervised Problems Type Attribute-Value. *Universidad y Sociedad*, 8(4), pp.43–53.

Rotondo, A. and Quilligan, F. (2020). Evolution Paths for Knowledge Discovery and Data Mining Process Models. *SN Computer Science*, 1(2). doi:https://doi.org/10.1007/s42979-020-0117-6.

Rowley, J. (2008). Understanding digital content marketing. *Journal of Marketing Management*, 24(5-6), pp.517–540. doi:https://doi.org/10.1362/026725708x325977.

Rumbul, R. (2016). ICTs, Openness and Citizen Perceptions of Government: How Civic Technologies Can Facilitate External Citizen Efficacy. *Medijske studije*, 7(14), pp.32–47. doi:https://doi.org/10.20901/ms.7.14.4.

Rustad, E. and Sæbø, Ø. (2013). *How, Why and with Whom Do Local Politicians Engage on Facebook?* [online] Semantic Scholar. doi:https://doi.org/10.1007/978-3-642-40346-0_7.

S, V. and R, J. (2016). Text Mining: open Source Tokenization Tools – An Analysis. *Advanced Computational Intelligence: An International Journal (ACII)*, 3(1), pp.37–47. doi:https://doi.org/10.5121/acii.2016.3104.

Salminen, J., Yoganathan, V., Corporan, J., Jansen, B.J. and Jung, S.-G. (2019). Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type. *Journal of Business Research*, 101(1), pp.203–217. doi:https://doi.org/10.1016/j.jbusres.2019.04.018.

Saltz, J. (2020). *CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects*. [online] Data Science Project Management. Available at: https://www.datascience-pm.com/crisp-dm-still-most-popular/.

Samsing, C. (2018). ¿Qué es Inbound Marketing? *Blog.hubspot.es*. [online] Available at: https://bit.ly/3pQq28S.

Sancho, R. (1990). Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. *Revista Española De Documentación Científica*, 13(3-4), pp.842–865. doi:https://doi.org/10.3989/redc.1990.v13.i3.842.

Santoso, A.D., Rinjany, D. and Bafadhal, O.M. (2020). Social Media and Local Government in Indonesia: Adoption, Use and Stakeholder Engagement. *Romanian Journal of Communication and Public Relations*, 22(3), p.21. doi:https://doi.org/10.21018/rjcpr.2020.3.307.

Saura, J.R. (2020). Using Data Sciences in Digital Marketing: Framework, methods, and Performance Metrics. *Journal of Innovation & Knowledge*, [online] 6(2). doi:https://doi.org/10.1016/j.jik.2020.08.001.

Saura, J.R., Ribeiro-Soriano, D. and Zegarra Saldaña, P. (2022). Exploring the challenges of remote work on Twitter users' sentiments: From digital technology development to a post-pandemic era. *Journal of Business Research*, [online] 142(1), pp.242–254.

doi:https://doi.org/10.1016/j.jbusres.2021.12.052.

Saxton, G.D., Niyirora, J.N., Guo, C. and Waters, R.D. (2015). #AdvocatingForChange: The Strategic Use of Hashtags in Social Media Advocacy. *Advances in Social Work*, [online] 16(1), pp.154–169. doi:https://doi.org/10.18060/17952.

Schreiner, M., Fischer, T. and Riedl, R. (2019). Impact of content characteristics and emotion on behavioral engagement in social media: literature review and research agenda. *Electronic Commerce Research*, 21(2), pp.329–345. doi:https://doi.org/10.1007/s10660-019-09353-8.

Schröer, C., Kruse, F. and Gómez, J.M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181(2), pp.526–534. doi:https://doi.org/10.1016/j.procs.2021.01.199.

Schultz, D. (2016). The future of advertising or whatever we're going to call it. *Journal of Advertising*, 45(3), pp.276–285.

Scott, D. (2020). *The new rules of marketing & PR*. 7th ed. New Jersey: John Wiley & Sons Inc.

Shah, D. and Murthi, B.P.S. (2020). Marketing in a data-driven Digital world: Implications for the Role and Scope of Marketing. *Journal of Business Research*, 125. doi:https://doi.org/10.1016/j.jbusres.2020.06.062.

Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A. and Chanona-Hernández, L. (2014). Syntactic N-grams as machine learning features for natural language processing. *Expert Systems with Applications*, 41(3), pp.853–860. doi:https://doi.org/10.1016/j.eswa.2013.08.015.

Siebers, V., Gradus, R. and Grotens, R. (2018). Citizen engagement and trust: A study among citizen panel members in three Dutch municipalities. *The Social Science Journal*, 56(4), pp.545–554. doi:https://doi.org/10.1016/j.soscij.2018.09.010.

Silverman, M. (2012). *Capturing Community: how to build, manage, and market your online community*. 1st ed. Cleveland, Ohio: CMI Books, p.14.

Skoric, M.M., Zhu, Q., Goh, D. and Pang, N. (2016). Social media and citizen engagement: A meta-analytic review. *New Media & Society*, 18(9), pp.1817–1839.

Song, C. and Lee, J. (2015). Citizens' Use of Social Media in Government, Perceived Transparency, and Trust in Government. *Public Performance & Management Review*, [online] 39(2), pp.430–453. doi:https://doi.org/10.1080/15309576.2015.1108798.

Sweeney, J. (2019). Public sector marketing pro. *JS press*, pp.95–96.

Taiminen, K. and Ranaweera, C. (2019). Fostering brand engagement and value-laden trusted B2B relationships through digital content marketing. *European Journal of Marketing*, pp.1759–1781. doi:https://doi.org/10.1108/ejm-10-2017-0794.

Thackeray, R., Neiger, B.L., Smith, A.K. and Van Wagenen, S.B. (2012). Adoption and use of social media among public health departments. *BMC Public Health*, 12(1). doi:https://doi.org/10.1186/1471-2458-12-242.

Torres, J.A.S., Cañada, F.J.A., Sandoval, A.V. and Alzate, J.A.S. (2021). Adoption of e-government in Colombia: the importance of government policy in citizens' use of e-government. *Electronic Government, an International Journal*, 17(2), p.220. doi:https://doi.org/10.1504/eg.2021.114577.

TWITTER (2022). *Política de Privacidad*. [online] twitter.com. Available at: https://twitter.com/es/privacy.

United Nations (2022). *E-Government Survey 2022 The Future of Digital Government*. [online] *United Nations*. Available at: https://desapublications.un.org/sites/default/files/publications/2022-09/Report%20without%20annexes.pdf.

Valenzuela, L., Merigó, J., Johnston, W., Nicolas, C. and Jaramillo, J. (2017). Thirty years of the Journal of Business & Industrial Marketing: a bibliometric analysis. *Journal of Business & Industrial Marketing*, 32(1), pp.1–17.

Van Osch, W. and Coursaris, C. (2014). Social Media Research: An Assessment of the Domain's Productivity and Intellectual Evolution. *Communication Monographs*, 81(3), pp.285–309.

Vinerean, S. (2017). Content Marketing Strategy. Definition, Objectives and Tactics. *Expert Journal of Marketing*, 5(2), pp.92–98.

Vinerean, S. and Opreana, A. (2021). Measuring Customer Engagement in Social Media Marketing: A Higher-Order Model. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(7), pp.2633–2654. doi:https://doi.org/10.3390/jtaer16070145.

Vivek, S.D., Beatty, S.E. and Morgan, R.M. (2012). Customer Engagement: Exploring Customer Relationships Beyond Purchase. *Journal of Marketing Theory and Practice*, 20(2), pp.122–146. doi:https://doi.org/10.2753/mtp1069-6679200201.

Vragov, R. and Kumar, N. (2013). The impact of information and communication technologies on the costs of democracy. *Electronic Commerce Research and Applications*, 12(6), pp.440–448. doi:https://doi.org/10.1016/j.elerap.2013.06.003.

Warren, A.M., Sulaiman, A. and Jaafar, N.I. (2014). Social media effects on fostering online civic engagement and building citizen trust and trust in institutions. *Government Information Quarterly*, 31(2), pp.291–301. doi:https://doi.org/10.1016/j.giq.2013.11.007.

Welby, B. (2019). The impact of digital government on citizen well-being. *OECD Working Papers on Public Governance*, 32. doi:https://doi.org/10.1787/24bac82f-en.

Wirth, R. and Chrysler, D. (2000). CRISP-DM: Towards a standard process model for data mining. In: *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*. pp.29–39.

Wirtz, B.W., Daiser, P. and Mermann, M. (2017). Social Media as a Leverage Strategy for Open Government: An Exploratory Study. *International Journal of Public Administration*, 41(8), pp.590–603.

doi:https://doi.org/10.1080/01900692.2017.1289388.

Wirtz, J., den Ambtman, A., Bloemer, J., Horváth, C., Ramaseshan, B., van de Klundert, J., Gurhan Canli, Z. and Kandampully, J. (2013). Managing brands and customer engagement in online brand communities. *Journal of Service Management*, 24(3), pp.223–244. doi:https://doi.org/10.1108/09564231311326978.

Wowczko, I. (2015). A Case Study of Evaluating Job Readiness with Data Mining Tools and CRISP-DM Methodology. *International Journal for Infonomics*, 8(3), pp.1066–1070. doi:https://doi.org/10.20533/iji.1742.4712.2015.0126.

Wuebben, J. (2011). *Content is Currency: Developing Powerful Content for Web and Mobile*. 1st ed. Boston: Nicholas Brealey Publishing, p.5.

Zhang, J. and Du, M. (2020). Utilization and effectiveness of social media message strategy: how B2B brands differ from B2C brands. *Journal of Business & Industrial Marketing*, 35(4), pp.721–740. doi:https://doi-org.sire.ub.edu/10.1108/JBIM-06-2018-0190.

Zheng, L. and Zheng, T. (2014). Innovation through social media in the public sector: Information and interactions. *Government Information Quarterly*, 31(0740-624X), pp.106–117. doi:https://doi.org/10.1016/j.giq.2014.01.011.