UNIVERSIDADE FEDERAL DO OESTE DO PARÁ
IEG-INSTITUTO DE ENGENHARIA E GEOCIÊNCIAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

PEDRO HENRIQUE COSTA MENEZES

VISÃO GERAL DO MERCADO EM SOCIAL CRM: UMA ANÁLISE DE ANÚNCIOS DE EMPREGO

SANTARÉM-PA
2024

PEDRO HENRIQUE COSTA MENEZES

VISÃO GERAL DO MERCADO EM SOCIAL CRM: UMA ANÁLISE DE ANÚNCIOS DE EMPREGO

Trabalho de Conclusão de Curso apresentado ao Programa de Computação, para obtenção do grau de Bacharel em Ciências da Computação; Universidade Federal do Oeste do Pará, Instituto de Engenharia e Geociências.

Orientador: Fábio Manoel França Lobato

SANTARÉM-PA

2024

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema Integrado de Bibliotecas – SIBI/UFOPA

M543v Menezes, Pedro Henrique Costa

Visão geral do mercado em social CRM: uma análise de anúncios de emprego./ Pedro Henrique Costa Menezes. - Santarém, 2024.

14 p. : il.

Inclui bibliografias.

Orientador: Fábio Manoel França Lobato.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Oeste do Pará, Instituto de Engenharia e Geociências, Bacharelado em Ciências da Computação.

1. Análise de dados. 2. Social CRM. 3. Desenvolvimento. 4.Análise de Vagas - Emprego. 5. Mineração de texto. 6. Análise de Mercado. I. Lobato, Fábio Manoel França, *orient*. II. Título.

CDD: 23 ed. 004.21

Bibliotecária - Documentalista: Cátia Alvarez – CRB/2 843

FORMULÁRIO DE AVALIAÇÃO DE TCC

Identificação:

Título do Trabalho:	Market Overview in Social CRM: An analysis of job advertisements
Aluno (a):	Pedro Henrique Costa Menezes
Orientador (a):	Fábio Manoel França Lobato

Avaliação:

Examinador (a) 1: Fábio Manoel França Lobato	**Nota:** 10
Assinatura:	

Examinador (a) 2: Marcio Jose Moutinho Da Ponte	**Nota:** 10,0
Assinatura:	

Examinador (a) 3: Marcelino Silva da Silva	**Nota:** 10,0
Assinatura:	

Parecer:

Resumo da Avaliação:

- [x] Aceitação incondicional
- [ ] Aceitação condicionada a modificações (especificar no verso)
- [ ] Recusado

Nota Final: 10,0

Santarém-PA, 8 de maio de 2024.

Presidente da Banca Examinadora

RESUMO

Os avanços tecnológicos e as constantes mudanças no mercado impulsionaram significativamente a evolução do Social Customer Relationship Management (Social CRM). À medida que as plataformas digitais e as mídias sociais ganham destaque, as empresas buscam estreitar o relacionamento com seus clientes, tornando essencial que os profissionais do setor acompanhem as tendências e se adaptem às novas demandas. Diante desse fenômeno, há uma alta frequência de ofertas de emprego relacionadas ao Social CRM. A avaliação destes anúncios constitui uma excelente oportunidade para acompanhar o ritmo acelerado do mercado de trabalho e identificar os elementos críticos para o sucesso nesta área. Neste contexto, este trabalho tem como objetivo coletar e analisar anúncios de emprego relacionados ao Social CRM e aplicar técnicas de mineração de texto para identificar os domínios de conhecimento e conjuntos de habilidades necessários aos profissionais desta área. Através desta abordagem, pretende-se desenvolver uma taxonomia que mapeie de forma sistemática e abrangente as competências exigidas neste mercado em constante mudança. Foram analisados 1.086 anúncios de emprego relacionados ao Social CRM. A análise forneceu uma visão geral detalhada das habilidades, tecnologias e conhecimentos necessários. Esta informação oferece boas oportunidades para empregadores e candidatos no processo de contratação e recrutamento, permitindo que ambos identifiquem com maior precisão as competências essenciais necessárias para se destacarem no setor do Social CRM. Além disso, os resultados desta pesquisa podem incentivar o desenvolvimento de estudos relacionados a esta área em expansão, contribuindo para o enriquecimento e aprimoramento do conhecimento em Social CRM.

Palavras-Chave: Análise de dados, Social CRM, Análise de Vagas de Emprego, Mineração de texto, Análise de Mercado.

ABSTRACT

Technological advances and constant changes in the market have significantly driven the evolution of Social Customer Relationship Management (Social CRM). As digital platforms and social media gain prominence, companies seek to strengthen their relationship with their customers, making it essential for industry professionals to keep up with trends and adapt to new demands. Given this phenomenon, there is a high frequency of job postings related to Social CRM. Evaluating these advertisements provides an excellent opportunity to keep up with the fast pace of the job market and identify the critical elements for success in this area. In this context, this work aims to collect and analyze job advertisements related to Social CRM and apply text-mining techniques to identify the knowledge domains and skill sets needed by professionals in this field. Through this approach, the intention is to develop a taxonomy that systematically and comprehensively maps the competencies required in this constantly changing market. We analyzed 1,086 job advertisements related to Social CRM. The analysis provided a detailed overview of the required skills, technologies, and knowledge. This information offers good opportunities for employers and candidates in the hiring and recruitment process, allowing both to identify more precisely the essential skills needed to excel in the Social CRM sector. In addition, the results of this research can encourage the further development of studies related to this expanding area, contributing to the enrichment and improvement of knowledge in Social CRM.

Keywords: Data Analysis, Social CRM, Job Analysis, Text Mining, Market Analysis

SUMÁRIO

2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) | 979-8-3503-0918-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/WI-IAT59888.2023.00092

Market Overview in Social CRM: An analysis of job advertisements

Pedro C. Menezes
Federal University of Western Pará
Santarém, PA, Brazil

Adrielson F. Justino
State University of Maranhão
São Luís, MA, Brazil

Barbara A. P. Barata
State University of Maranhão
São Luís, MA, Brazil

Antonio F. L. Jacob Junior
State University of Maranhão
São Luís, MA, Brazil
antoniojunior@professor.uema.br

Fábio M. F. Lobato
Federal University of Western Pará
Santarém, PA, Brazil
fabio.lobato@ufopa.edu.br

***Abstract*—Technological advances and constant changes in the market have significantly driven the evolution of Social Customer Relationship Management (Social CRM). As digital platforms and social media gain prominence, companies seek to strengthen their relationship with their customers, making it essential for industry professionals to keep up with trends and adapt to new demands. Given this phenomenon, there is a high frequency of job postings related to Social CRM. Evaluating these advertisements provides an excellent opportunity to keep up with the fast pace of the job market and identify the critical elements for success in this area. In this context, this work aims to collect and analyze job advertisements related to Social CRM and apply text-mining techniques to identify the knowledge domains and skill sets needed by professionals in this field. Through this approach, the intention is to develop a taxonomy that systematically and comprehensively maps the competencies required in this constantly changing market. We analyzed 1,086 job advertisements related to Social CRM. The analysis provided a detailed overview of the required skills, technologies, and knowledge. This information offers good opportunities for employers and candidates in the hiring and recruitment process, allowing both to identify more precisely the essential skills needed to excel in the Social CRM sector. In addition, the results of this research can encourage the further development of studies related to this expanding area, contributing to the enrichment and improvement of knowledge in Social CRM.**

***Index Terms*—Data Analysis, Social CRM, Job Analysis, Text Mining, Market Analysis**

I. Introduction

Companies are looking to strengthen customer relationships with the continuous expansion of digital platforms and social media usage [1], [2]. The Social Customer Relationship Management (Social CRM) concept has been significantly boosted, making it an area of extreme relevance in the current scenario [3]. The increased use of the Internet, particularly social networking platforms such as Facebook, Twitter, Instagram, and Snapchat, has brought several opportunities and challenges for marketers looking to manage customer relationships [4]. Therefore, professionals in this sector need to follow trends and adapt to new demands. In this scenario, Customer Relationship Management (CRM) has gradually evolved to take advantage of new technologies and emerging channels, moving from database-driven CRM to electronic CRM (eCRM), mobile CRM (mCRM), and later Social CRM, the last integrates social media tools and customer management to monitor market analysis.

Social CRM emphasizes active participation in online conversations and communities, enabling companies to monitor and respond to social interactions such as comments, mentions, ratings, and shares [3]. In this way, it is possible to build stronger customer relationships and gain valuable insights into their preferences and opinions [1], [5]. Faced with this phenomenon, a growing number of companies have adopted online recruitment as a way to attract candidates [6]. So, using the internet to publicize job postings offers an excellent opportunity to monitor the job market in real-time [7]. A job vacancy consists of two main elements: a title summarizing the position and a detailed description that includes the job requirements and skills needed by the candidate [7].

Text mining in textual databases has become essential to support knowledge discovery in an agile and cost-effective way, surpassing traditional sources [8]. Applying text mining methods to analyze job posts allows for understanding employers' needs, and identifies skills, occupational changes, and working conditions [9]. Based on these facts, this study aimed to perform a textual content analysis on online job advertisements related to Social CRM to identify the knowledge domains, job requirements, skill sets, and tools needed in this segment.

The remainder of the article is organized as follows. In Section II related works are presented. Section III describes the research process. Section IV presents the results resulting from the analyses. Finally, Section V presents the final considerations.

II. Related Work

[10] focused on proposing a semi-autonomous methodology to analyze job advertisements in software engineering and big data. The authors used the Indeed platform as the data source. Furthermore, Latent Dirichlet Allocation (LDA) was chosen as the most appropriate topic modeling method

979-8-3503-0918-8/23/$31.00 ©2023 IEEE
DOI 10.1109/WI-IAT59888.2023.00092

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.

since this method considers the semantics of advertisements, revealing the knowledge domain and critical skills needed for the job market. In addition, the method used can be improved to automatically analyze ads' text and routinely update the skills list.

[11] aimed to understand the most requested skills in big data job advertisements, considering aspects related to the location (city), salary, educational background, and experience. The data were extracted from a Chinese website called Zhaopin, and the texts were grouped using the k-means algorithm. The work highlights the need to integrate more text mining technologies to explore further the information and conditions of the big data job market.

[12] focused on comparing keyword dictionary methods and the LDA model for content analysis in job advertisements. The authors found that the topic identification algorithm performed better when compared to the approach based on the European Dictionary of Skills and Competencies (DISCO). However, there are disadvantages to adopting the LDA model, such as identifying topics that are difficult to interpret. The authors emphasize the preciousness of these discoveries as a way to support communication between the educational network and the labor market.

[13] sought a broader understanding of the requirements related to the spheres of Information Technology in the Commonwealth of Independent States region by mapping the most accurate combinations of soft and hard skills. The methodology proposed in the work is based on natural language processing, hierarchical grouping, and association mining, showing that the methods used can be put in a theoretical and practical way.

The related works have shown how techniques such as N-gram, topic modeling, clustering, and natural language processing are pertinent to reveal important information in many fields. In addition, the works show the importance of the information obtained by analyzing the data in the alignment of education and the labor market. It is essential to point out that a limitation is noted regarding the analysis of online job openings in Social CRM about the needs of the labor market. Thus, the differential of this work consists of a pioneering and comprehensive approach in the analysis of job openings, highlighting the job market's main requirements in the Social CRM field.

III. Materials and Methods

The research method adopted in this work is the Cross-Industry Standard Process for Data Mining (CRISP-DM) used in [14]. This process model is well established for conducting data mining projects [15]. The application of this model is divided into six stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Implementation, as shown in Figure 1.

A. Business understanding

In the first stage, the business or application domain was understood. For this, it is necessary to understand the project's goals and identify the research problems based on data mining. It is noteworthy that the definition of the research goals in this study was carried out together with a literature review to obtain information on the role of professionals, techniques, and procedures for mapping the knowledge domains and skills required in the market in Social CRM, according to the related works.

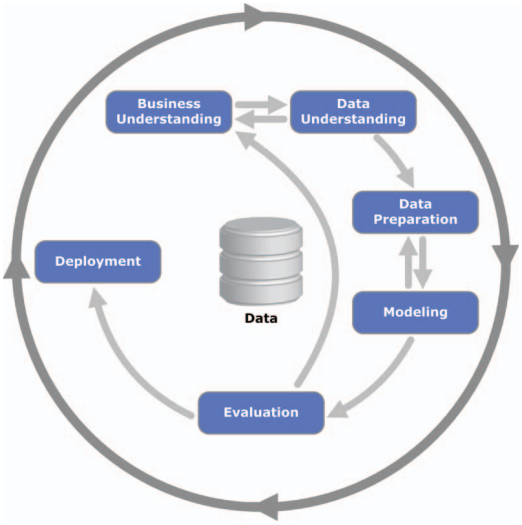

Fig. 1. Phases of the CRISP-DM method. Adapted from [16].

At this stage, an initial data collection was carried out to familiarize them with their characteristics, identifying problems and insights about the data. Poorly designed data is challenging to extract and use on the web in an automated process [17]. In this sense, we use Diffbot, a web scraping tool that can easily collect data from web pages, convert it into some desired format, and store it in some local repository. Through it, it is possible to collect information from various sources on the internet, including news, articles, products, social media, and job vacancies. The platform automatically identifies patterns in the data, such as titles, descriptions, and other relevant elements.

B. Data understanding

We collected 1,296 job advertisements from various websites related to job opportunities using the keywords *"Social Customer Relationship Management"*, *"Social CRM"*, *"eCRM"*, *"mCRM"* and *"iCRM"*. Data were downloaded in Comma-Separated Values (CSV) file format. To understand the composition of the database, several aspects were evaluated, including the data volume, and missing data, to obtain a precise notion of its dimensionality. Table I displays the fields present in the database, including *id*, *title*, *text*, *humanlanguage*, *Uniform Resource Locator (URL)*(URL), *requirements* and *tasks*. It is noteworthy that of the 1,296 advertisements collected, some did not contain information in the fields *title*, *requirements*, and *tasks*. Based on the field *text*, it was possible to determine that the average text length is 758 words.

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COLLECTED DATASET FIELDS AND THEIR INFORMATION.

Fields	Description	Amount
Id	Identifier number	1,296
Title	Job title from the corresponding job advertisement	1,191
Text	Job description body text	1,296
Humanlanguage	Text Language	1,296
PageUrl	Job posting source URL	1,296
Requirements	Information related to requirements	731
Tasks	Information about job duties	977

Incorrect data, such as duplicates and *"no description available"* have been removed. Then, based on the average text length of the job descriptions, those containing more than 1,500 words were removed. It was performed because *outliers* were identified, such as content not related to the jobs description (*e.g.,* companies history, website information *etc*). The final dataset included 1,086 documents.

C. Data preparation

From the data understanding, the advertisement titles and text attributes were selected as the main features to be used in the modeling phase. The preprocessing and manipulation steps of the textual content applied to the data were performed using the conventional text mining pipeline [18], [19]. The textual content went through the following preprocessing steps: *a) Removal of duplicates:* The PageUrl field was secondarily manipulated and consisted of removing the prefix and suffixes from the URLs, so, removing duplicate ads were based on checking the identifier (id) and page address (PageUrl) repeated to ensure that each ad appeared only once in the dataset; *b) Text standardization:* The text was converted to lowercase letters to avoid discrepancies due to differences in capital letters, non-English language ads have also been removed; *c) Removal of accents, punctuation, irrelevant characters, URLs, and excessive spaces:* These steps were applied to eliminate elements that do not contribute to the analysis; d) Removal of stopwords: The stopwords from the Natural Language Toolkit (NLTK) library were used - this allowed access to a predefined list of English stopwords removed from the text. In addition to using the NLTK set of stopwords, other words were manually added to the stopwords list. Those added words included: *"view"*, *"e g"*, *"please"*, *"send"*, *"end"*, *"de"*, *"compliance"*, *"risk"*, *"risks"*, *"use"*, *"sanctions"*, *"france"*, *"united"*, *"kingdom"*, *"york"*, *"philippinesofficer"*, *"america"*, *"abbott"*, *"nbsp"*, *"ireland"*, *"northern"*, *"belfast"*, *"citi"*, *"chalhoub"*, *"bcg"*, *"per"*, *"hq"*, *"rohq"*, *"fiserv"*, *"aml"*, *"surveillance"*, *"surveillances"*, *"ct"*, *"legal"*, *"regulatory"*, *"regulations"*, *"rules"*, *"law"*, *"hybridbelfast"*, *"citiover"*, *"tts"*, *"grct"*, *"gft"*, *"email"*, *"apply"*, *"career"*; Finally, *e) Removal of numerical characters:* This step was applied only after evaluating the years of experience required to obtain more accurate results for the topic modelling.

D. Modeling

Several data mining techniques were explored to identify patterns in job advertisements. One approach used to analyze continuous sequences of words was the generation of N-grams, which are neighboring sequences of items in a document [20]. Topic modeling was another technique employed to extract knowledge from unstructured texts in job descriptions. This method aims to group documents and words with similar meanings and is widely used in NLP [21]. Among the various topic modeling options available, for example, Latent Dirichlet Allocation, Non-Negative Matrix Factorization (NMF), BERTopic, Word2Vec [22]–[24]. BERTopic was chosen since it considers semantic information. This approach generates document representations through embeddings of language models based on pre-trained transformers, and a class-based TF-IDF to create dense clusters, allowing the creation of easily interpretable topics and highlighting essential words in topic descriptions [24].

Unlike other techniques such as LDA, BERTopic does not require prior knowledge of the number of clusters [24]. However, it can generate a large number of clusters, making semantic interpretation difficult. To deal with this, it is possible to reduce the number of topics after training, aiming to create a realistic and easily annotated amount by human experts, following the human-in-the-loop approach [25]. In addition, BERTopic offers a *reduce_topics* method that uses the original template information to do a topic reduction, regrouping, and combining existing topics to form broader topics [24]. This can be useful to simplify the interpretation of the results, thus generating meaningful insights.

It should be noted that the original model trained by BERTopic contains Topic -1, which always represents outliers and should not be considered in the analyses [24], [26]. Preprocessing is generally unnecessary for BERTopic as it relies on an embedding approach, and keeping the original structure of the text is vital for transformer models [26]. However, in cases of noisy data, as in the context of this work, conducting preprocessing steps can become beneficial. Refining and cleaning the data before applying BERTopic substantially improved the obtained results.

The third approach was related to building the lexicon related to job posts. The development of this lexicon involved a semi-automatic approach, as described in Figure 2. Initially, the axes were defined, considering the existing elements in the advertisement structure, which resulted in job posts, technologies, skills (technical and behavioral), and knowledge. Next, the authors read a representative set of job postings (around 30% of the dataset was carefully examined).

This step was combined with a script for searching terms related to keywords observed during the reading, which are: *"platforms"*, *"platform"*, *"tools"*, *"software"*, *"knowledge"*, *"certification"*, *"plus"*, *"ability"*, *"required"*, *"programs"* e *"experience"*. This process culminated in building a dictionary that facilitated identifying and visualizing the topics addressed in the job posts. Finally, the terms added to the dictionary were

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.

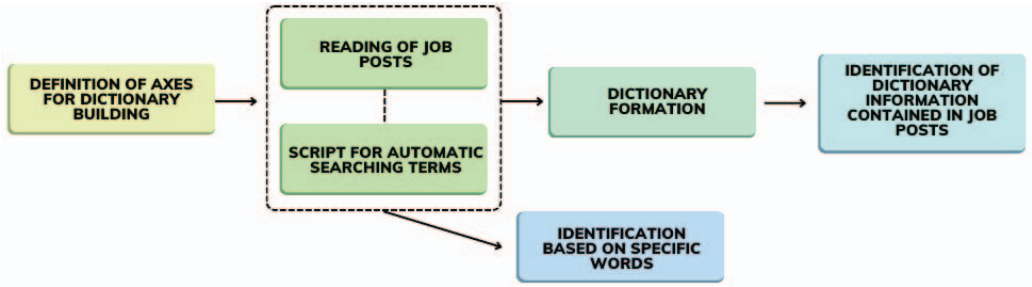

Fig. 2. Stages for building the job posting lexicon.

used to determine the most recurrent ones in the job posts. It is worth mentioning that the evaluation of the results was carried out later, which is described in the following subsection.

E. Evaluation and implementation

The evaluation stage was based on the grounded theory approach, which consists of a qualitative research method that allows researchers to discern explicit and implicit processes in their data [27]. In this study, the approach was used so that the authors could read a representative data sample and cross them with the results obtained. It is also worth mentioning that a sample calculation was made, considering the population's size, a confidence level of 95%, and a margin of error of 5.

The sample size was therefore set at 297 job advertisements, chosen at random, which were read in full as a complementary way of understanding job post characteristics. The authors analyzed the job advertisements to identify domain terms related to Social CRM in order to build the lexicon and increase the study reliability. The dataset with 1,086 job advertisements was used for N-gram analysis, topic modeling, and term counting from the lexicon built.

F. Technologies

The Python (3.9.13) was used to develop the scripts. The experiments were conducted in Jupyter Notebook (6.4.12). In the data preparation and modeling step, several libraries were used to perform different tasks. For example, Pandas (1.4.4) was used for data manipulation; the Natural Language Toolkit (MLTK)(3.7) was used for text preprocessing; Regular expressions were implemented using library (RE) (2022.7.9). The textual representation was performed using the Bag-of-Words method, implemented through the Scikit-learn library (1.0.2). For topic modeling, BerTopic[1] (0.15.0) was used. For data visualization, matplotlib/pyplot (3.5.2) and Seaborn(0.11.2) were adopted for generating graphics and WordCloud[2] (1.8.2.2).

IV. Results

After preprocessing the data, a total set of 1,086 job advertisements was analyzed. Figure 3 visually represents job descriptions in a . This representation provides insight into the skills, knowledge, and requirements referenced in available opportunities. Analyzing these terms highlights the importance of identifying similar words for a deeper understanding of job vacancies. By grouping related terms, it is possible to obtain a comprehensive view of the requirements demanded by employers in Social CRM.

By deepening our analysis of Figure 3, we identified aspects not previously explored in the research questions, highlighting the community's concern with inclusion and equity. This analysis reveals the presence of expressions such as *"equal opportunity"*, *"sexual orientation"*, *"gender identity"*, *"race color"* and *"religion"*. Such terms indicate the value employers give to diversity and guarantee equal opportunities for all candidates. Including these terms call attention to the importance of understanding the technical skills required for positions and the relevance of policies and values that guide the work environment. Furthermore, incorporating the words *"support"* and *"employer qualified"* suggests that companies are focused on creating a professional background in which all employees feel supported and recognized.

Analyzing job advertisements made it possible to examine the distribution of years of experience requested by companies, as shown in Figure 4. Figure 4 depicts an overview of minimum years of experience. Notably, it can be seen that the years of experience distribution vary. Most postings, corresponding to about 50, require candidates with more than 10 years of experience, indicating the search for highly specialized professionals with a broad professional background. The analysis also reveals a substantial group of jobs between 5 and 6 years old, meaning that many opportunities are destined for professionals with intermediate experience. It is important to note that the number of people with 0 to 2 years of experience is relatively low.

To better understand the skills in Social CRM. Initially, an N-gram analysis was carried out, in which the main opportunities and requirements were identified. Subsequently, these skill sets were mapped based on topic modeling. In addition, the most sought-after tools in Marketing, Design, Data Analysis, Software Development, and Programming Languages were analyzed. The results are presented and discussed below.

The preliminary results related to the N-grams in the titles are presented in Figure 5, in which it is possible to observe the 15 most frequent N-grams. The terms indicated the ever-

[1]https://maartengr.github.io/BERTopic/index.html
[2]https://pypi.org/project/wordcloud/

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.

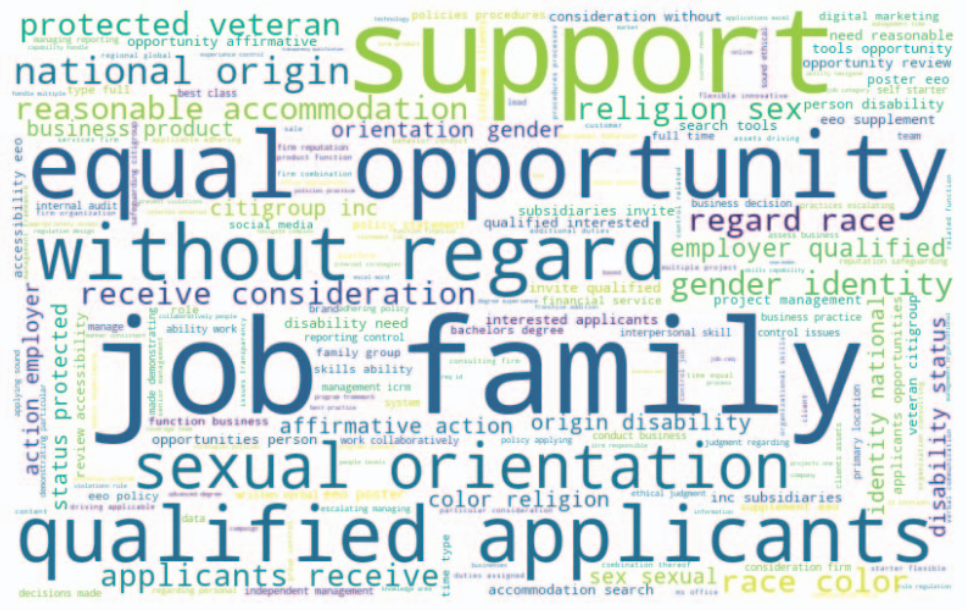

Fig. 3. Word cloud representation of job posts descriptions.

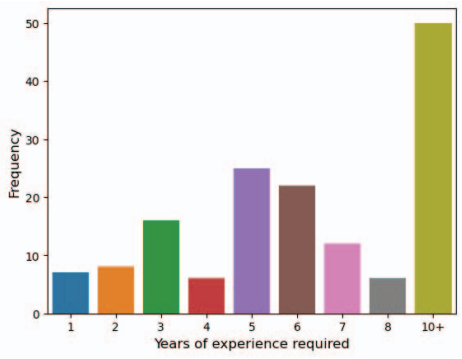

Fig. 4. Years of experience required.

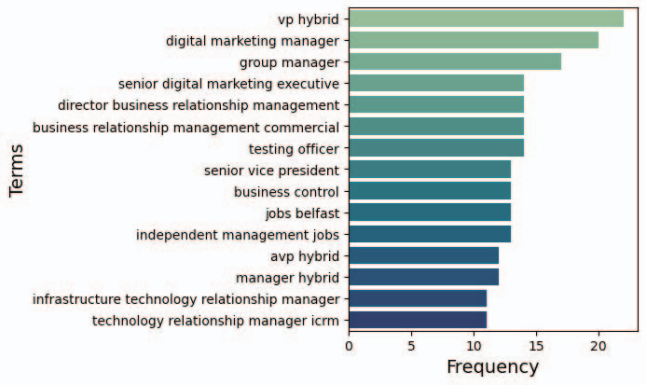

Fig. 5. Representation of N-grams constructed based on job titles.

evolving opportunities and demands related to Social CRM, reflecting the importance of technical skills, advanced research, and interdisciplinary collaboration. Through this analysis, it was possible to verify which titles appear in the job posts, fundamental to understanding which positions are being demanded.

Analyzing Figure 5, it is possible to identify some terms that are directly related to Social CRM, such as *"digital marketing manager", "business relationship management commercial", "senior digital marketing executive", and "director business relationship management"*. Based on a cross-analysis with Figure 4, there is a notable congruence since the high presence of senior and managerial positions corroborates the job vacancies that require professionals with extensive experience since these positions require a broader business view. In addition, these terms have keywords associated with customer relationship management and digital marketing, which are Social CRM essential elements.

On the other hand, some N-grams may not be directly related to Social CRM, such as *"jobs belfast"*, *"independent management jobs"* and *"business control"*. These terms can be considered noise and should be filtered out to ensure the search is more accurate and focused on relevant topics.

The BERTopic-based topic modeling analysis was performed on the job advertisements using different parameters, and, as expected, the initial model identified a large number of topics (31 topics), some of which were redundant or too specific. In this way, reducing topics can lead to a more concise and representative set, facilitating the analysis and understanding of the themes present in the documents. These topics are presented in Table II, along with the BERTopic descriptive

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.

keywords. Topics are listed in the table in descending order of frequency. Discovered topic names were assigned manually, considering descriptive keywords and their frequencies.

In Table II, the first word was the most seen, and the last was the least seen in a topic. In this context, topic names were commonly assigned, taking into account the first four keywords. On the other hand, for some topics, the keywords together defined a specific domain, as detailed below:

Topic 1: Business Management and Control - This topic indicates that companies may be looking for professionals with data management and analysis skills. The social CRM area might be related to managing customer data, analyzing marketing metrics, and implementing efficient processes to improve customer relationships. In addition, companies seek to comply with Equal Employment Opportunity (EEO) standards.

Topic 2: Social CRM and Marketing - This topic is directly relevant to Social CRM. The job posts indicate that companies are looking for professionals who can lead and execute digital marketing strategies and manage customer relationships through social media platforms and CRM systems. This includes campaign management skills, customer data analysis, and utilizing social media to drive customer engagement and loyalty.

Topic 3: Career Analyst and Diversity Officer - This topic addresses career management and promoting diversity in organizations. While not directly tied to Social CRM, there may be some linkage implicitly, as a diverse Social CRM team can better understand and address specific customers' needs. In addition, career analysis can help companies identify suitable talent to work with Social CRM and improve the teams' efficiency.

Topic 4: Business and Portfolio Strategy - This topic addresses business strategies and portfolio management issues, focusing on business expansion and growth. Although business strategy can influence lead capturing and collaborative development of products/services, this topic is not directly related to the specific practices and Social CRM tools.

Topic 5: Growth Ventures and Digital Marketing - This topic is related to the usage of digital marketing strategies to drive business growth and expansion, as well as looking for business opportunities that can benefit from using Social CRM to improve relationships with customers, and boost sales.

Topic 6: Partnership and Microsoft Campaign - This topic suggests the importance of strategic partnerships and campaigns with companies like Microsoft. Under Social CRM, this can involve integrating CRM solutions with Microsoft platforms/tools to improve efficiency, share data, and enhance the customer experience.

In general, topics 1, 2, 5, and 6 seem more in line with the field of Social CRM, as they address aspects related to business management, digital marketing, business growth, strategic partnerships, and campaigns. They can directly influence the Social CRM implementation. While topics 3 and 4 are relevant in other business areas, they do not directly relate to Social CRM.

Through N-gram analysis and topic modeling, it was possible to glimpse the diversity and nuances of professional opportunities in this field. However, the above reviews provide an overview of requirements and areas of focus, ranging from business management and digital marketing to data analysis. In this way, a lexicon was built to categorize and organize this complex information to obtain a more precise and structured view of the employment scenario in Social CRM.

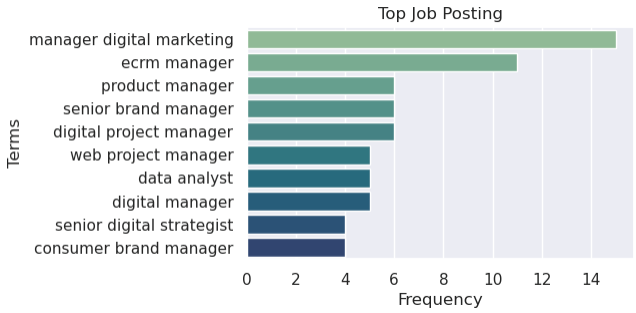

Fig. 6. Job Posting

Figure 6 shows the term frequency in the lexicon built from job posts related to Social CRM. The predominance of positions such as *"digital marketing manager"*, *"ecrm manager"*, and *"product manager"* suggests a growing need for professionals capable of leading digital and CRM strategies. This indicates a strong emphasis on digital marketing management and utilizing customer relationship systems to drive business results. The priority given to *"senior brand manager"*, *"consumer brand manager"*, *"digital project manager"* and *"web project manager"* reinforces CRM strategies relevance to build and maintain solid brands in the digital environment. The growing importance of data analysis is represented by the *"Data Analyst"*, suggesting the need for professionals capable of getting valuable insights from customer data.

Figure 7 reveals insights into key technologies present in job posts. Design tools such as *"photoshop"* and *"InDesign"* are prevalent, evidencing the market need for professionals capable of creating attractive and engaging visual content. Technologies related to data analysis, such as *"google analytics"*, *"tableau"*, and *"power bi"* indicate an increasing focus on data-driven decision-making.

Regarding the skills mentioned in the job posts through Figure 8, we can observe that the frequency of data analysis and programming skills calls attention to the need for professionals who can interpret valuable insights from the collected data and implement solutions techniques. The emphasis on *"communication"*, *"interpersonal skills"*, and *"flexible"* reflects the significance of interpersonal skills in a scenario where collaboration and adaptation are crucial to success. The prevalence of *"proactive"* suggests that this ability is highly valued in dealing with the ever-evolving complexities of the digital environment.

Figure 9 shows an overview of the main knowledge required by companies. The inclusion of *"digital marketing"*,

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.

TABLE II
REPRESENTATION OF SELECTED TOPICS WITH TERMS.

Topic	Labels	Key words
1	Business Management and Control	business, management, control, experience, data, related, internal, eeo, work, job
2	Social CRM and Marketing	marketing, digital, customer, crm, social, sales, media, brand, campaigns, management
3	Career Analyst and Diversity Officer	learn, analyst, career, diversity, jobs, officer, hybridmultiple, search, cookie, preferences
4	Business and Portfolio Strategy	business, commercial, portfolio, nutrition, strategy, management, global, divisional, brm, life
5	Growth Ventures and Digital Marketing	growth, ventures, architects, execute, customer, validate, digital, worlds, adoption, marketing
6	Partnership and Microsoft Campaign	marketing, partnership, drive, campaign, microsoft, windows, partners, consumer, team, across

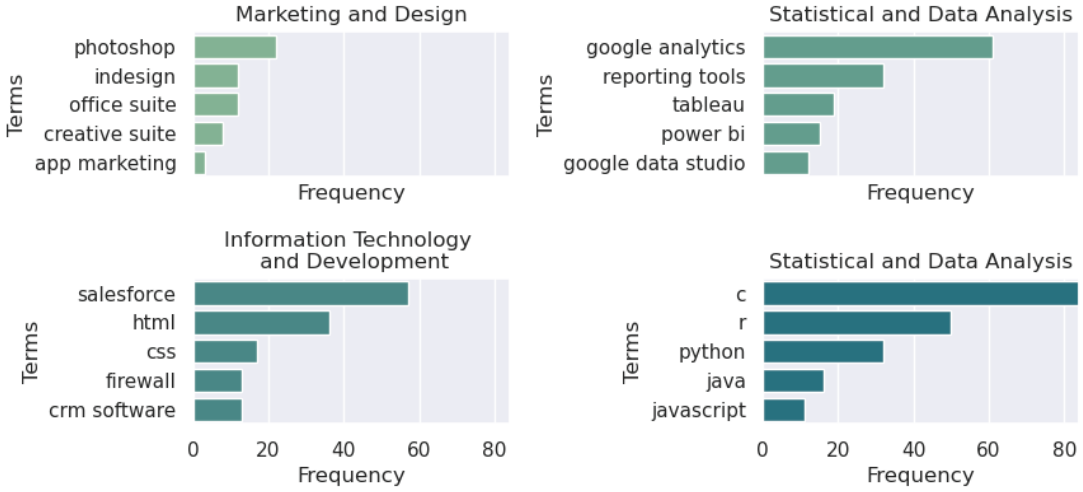

Fig. 7. Top Technologies

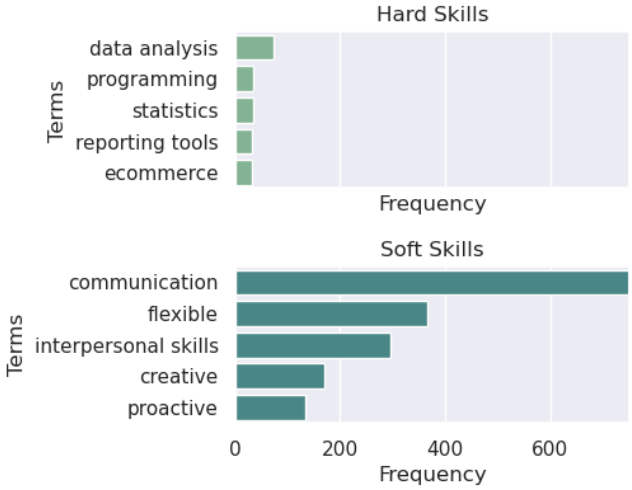

Fig. 8. Top Skills

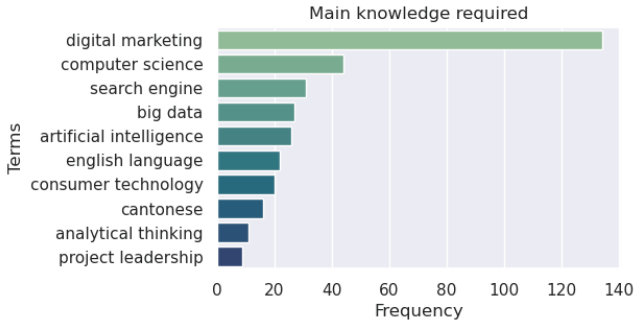

Fig. 9. Main knowledge required

"computer science", and *"big data"* as frequent terms show up the need for professionals with a solid understanding of modern digital marketing concepts/technologies. The emphasis on *"search engine"* and *"artificial intelligence"* suggests that optimizing for search engines and adopting artificial intelligence solutions are considered vital skills. The presence of the *"English language"* as essential knowledge reflects the globalization of the Social CRM field, thus, the importance of communication in an international context.

The results obtained from the lexicon indicate that Social CRM is rapidly evolving towards more sophisticated digital strategies and data-driven analytics. Professionals with a balance of technical and soft skills, whith a deep understanding of digital and technology trends, have a competitive advantage.

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.

V. Final Remarks

The research novelty relies on a comprehensive approach to analyzing job advertisements related to Social CRM by combining text mining and human-in-the-loop methods/strategies. To carry out the study, 1,086 job posts were collected using Diffbot, which facilitates data scraping. The analysis focused on identifying and distinguishing skills, knowledge, positions, and technologies. Good data science practices were used, which allowed the data analysis centered on the N-gram approach, topic modeling, and the building of a lexicon oriented to the Social CRM domain. In addition, the concepts related to grounded theory are highlighted to validate the results. These steps were crucial to understanding the Social CRM market's dynamics, patterns, and trends.

In short, the results converge to identify a growing demand for highly specialized professionals in Social CRM, who would be capable of leading digital and CRM strategies. However, we also identified non-relevant N-grams, suggesting the need to filter out noise to target the search more precisely. Topic modeling also suggests a growing demand for professionals with skills in digital strategies, data analysis, and customer relationship management.

This information has the potential to guide people looking to enter this field. It also makes promoting public policies aimed at this job market easier. Text mining and data analysis have become powerful tools for understanding the job market dynamics and making important decisions about hiring and qualifying professionals in this ever-changing field. In future work, we intend to expand our data collection, including other search terms. We also would like to refine our lexicon by inviting Human Resources specialists to validate the dictionary. Finally, we intend to apply more text mining methods, such as k-means for grouping the data, and large language models (e.g., GPT) for testing a question-and-answering approach.

Acknowledgements

This work was supported by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) - DT-308334/2020; and by *Fundação Amazônia de Amparo a Estudos e Pesquisas* (FAPESPA) PRONEM-FAPESPA/CNPq nº 045/202.

References

[1] M. M. Mariani, R. Perez-Vega, and J. Wirtz, "Ai in marketing, consumer research and psychology: A systematic literature review and research agenda," *Psychology & Marketing*, vol. 39, no. 4, pp. 755–776, 2022.

[2] F. Lobato, M. Pinheiro, A. Jacob, O. Reinhold, and Á. Santana, "Social crm: Biggest challenges to make it work in the real world," in *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers 19*. Springer, 2017.

[3] R. Alt and O. Reinhold, *Social customer relationship management: Fundamentals, applications, technologies*. Springer Nature, 2019.

[4] W. Silva, Á. Santana, F. Lobato, and M. Pinheiro, "A methodology for community detection in twitter," in *Proceedings of the International Conference on Web Intelligence*, 2017, pp. 1006–1009.

[5] B. N. R. Chagas, J. A. N. Viana, O. Reinhold, F. Lobato, A. F. Jacob, and R. Alt, "Current applications of machine learning techniques in crm: a literature review and practical implications," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 2018.

[6] A. B. Holm, "The effect of e-recruitment on the recruitment process: Evidence from case studies of three danish mncs," in *Proceedings of the 3rd European academic workshop on electronic human resource management*, 2010, pp. 91–111.

[7] R. Boselli, M. Cesarini, S. Marrara, F. Mercorio, M. Mezzanzanica, G. Pasi, and M. Viviani, "Wolmis: A labor market intelligence system for classifying web job vacancies," *Journal of intelligent information systems*, vol. 51, pp. 477–502, 2018.

[8] K. Ghazzawi and A. Accoumeh, "Critical success factors of the e-recruitment system," *Journal of Human Resources Management and Labor Studies*, vol. 2, no. 2, pp. 159–170, 2014.

[9] L. M. Kureková, M. Beblavỳ, and A. Thum-Thysen, "Using online vacancies and web surveys to analyse the labour market: A methodological inquiry," *IZA Journal of Labor Economics*, vol. 4, pp. 1–20, 2015.

[10] F. Gurcan and N. E. Cagiltay, "Big data software engineering: Analysis of knowledge domains and skill sets using lda-based topic modeling," *IEEE access*, vol. 7, pp. 82 541–82 552, 2019.

[11] D. Debao, M. Yinxia, and Z. Min, "Analysis of big data job requirements based on k-means text clustering in china," *PloS one*, 2021.

[12] Z. Ao, G. Horvath, C. Sheng, Y. Song, and Y. Sun, "Skill requirements in job advertisements: A comparison of skill-categorization methods based on explanatory power in wage regressions," *arXiv preprint arXiv:2207.12834*, 2022.

[13] A. Ternikov, "Soft and hard skills identification: insights from it job advertisements in the cis region," *PeerJ Computer Science*, vol. 8, 2022.

[14] S. Moro, R. Laureano, and P. Cortez, "Using data mining for bank direct marketing: An application of the crisp-dm methodology," in *Proceedings of European Simulation and Modelling Conference - ESM2011*, 2011.

[15] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, vol. 1. Manchester, 2000, pp. 29–39.

[16] K. Jensen, "A diagram showing the relationship between the different phases of crisp-dm and illustrates the recursive nature of a data mining project," *Wikimedia Commons*, vol. 26, 2012.

[17] H. Nigam and P. Biswas, "Web scraping: from tools to related legislation and implementation using python," in *Innovative Data Communication Technologies and Application: ICIDCA 2020*. Springer, 2021.

[18] D. Cirqueira, M. F. Pinheiro, A. Jacob, F. Lobato, and Á. Santana, "A literature review in preprocessing for sentiment analysis for brazilian portuguese social media," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 2018, pp. 746–749.

[19] L. C. Fernandes, J. Silva, A. Jacob, and F. Lobato, "An extensive analysis of online restaurant reviews: a case study of the amazonian culinary tourism," in *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2020, pp. 81–84.

[20] A. Balfagih, V. Keselj, and S. Taylor, "N-gram and word2vec feature engineering approaches for spam recognition on some influential twitter topics in saudi arabia," in *Proceedings of the 6th International Conference on Information System and Data Mining*, 2022, pp. 101–107.

[21] J. Garcia, G. Villavicencio, F. Altimiras, B. Crawford, R. Soto, V. Minatogawa, M. Franco, D. Martínez-Muñoz, and V. Yepes, "Machine learning techniques applied to construction: A hybrid bibliometric analysis of advances and future directions," *Automation in Construction*, vol. 142, p. 104532, 2022.

[22] D. Maier, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Häussler *et al.*, "Applying lda topic modeling in communication research: Toward a valid and reliable methodology," in *Computational methods for communication science*. Routledge, 2021.

[23] Y. Chen, H. Zhang, R. Liu, Z. Ye, and J. Lin, "Experimental explorations on short text topic mining between lda and nmf based schemes," *Knowledge-Based Systems*, vol. 163, pp. 1–13, 2019.

[24] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[25] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," *Future Generation Computer Systems*, vol. 135, pp. 364–381, 2022.

[26] R. Egger and J. Yu, "A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts," *Frontiers in sociology*, vol. 7, p. 886498, 2022.

[27] K. Charmaz and R. Thornberg, "The pursuit of quality in grounded theory," *Qualitative research in psychology*, vol. 18, no. 3, 2021.

Authorized licensed use limited to: UNIVERSIDADE DE SAO PAULO. Downloaded on April 06,2024 at 14:28:07 UTC from IEEE Xplore. Restrictions apply.