



UNIVERSIDADE
ESTADUAL DO
MARANHÃO

UNIVERSIDADE ESTADUAL DO MARANHÃO

CENTRO DE CIÊNCIAS TECNOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO E
SISTEMAS

MESTRADO PROFISSIONAL EM ENGENHARIA DE COMPUTAÇÃO E SISTEMAS

GUSTAVO NOGUEIRA DE SOUSA

ANÁLISE INTELIGENTE DE MÍDIAS SOCIAIS PARA
POTENCIALIZAR GESTÃO DO RELACIONAMENTO COM CLIENTES

São Luís

2021

UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO E
SISTEMAS
MESTRADO PROFISSIONAL EM ENGENHARIA DE COMPUTAÇÃO E SISTEMAS

GUSTAVO NOGUEIRA DE SOUSA

ANÁLISE INTELIGENTE DE MÍDIAS SOCIAIS PARA
POTENCIALIZAR GESTÃO DO RELACIONAMENTO COM CLIENTES

Dissertação de Mestrado apresentada ao Pro-
grama de Pós-Graduação em Engenharia de
Computação, Centro de Ciências Tecnológicas,
Universidade Estadual do Maranhão.
Orientador: Prof. Dr. Fábio Manoel Braga
Lobato.
Coorientador: Prof. M.Sc. Antônio Fernando
Loureiro Jacob Jr.

São Luís

2021

Sousa, Gustavo Nogueira de.
Análise insipiente de mídias locais para potencializar gestão de relacionamento com clientes / Gustavo Nogueira de Sousa. – São Luís, 2021.

124f.

Dissertação (Mestrado) – Curso de Engenharia de Computação e Sistemas, Universidade Estadual do Maranhão, 2021.

Orientador: Prof. Dr. Fábio Manoel Pimenta Lobato.

1. Mídias locais. 2. Mineração de texto. 3. Aprendizagem de máquina. 4. Gestão de relacionamento com clientes. 5. Social CRM. I. Título.

CDU 004.738.3-895.26

Elaborado por Giselle Frazão Tavares - CRB 13/663

Gustavo Nogueira de Sousa

Análise inteligente de Mídias Sociais para potencializar gestão do relacionamento com clientes

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Computação e Sistemas da Universidade Estadual do Maranhão, como parte das exigências para a obtenção do título de Mestre em Engenharia de Computação e Sistemas.

Tribunal aprovado, São Luís, 09 de Setembro de 2021.



Prof. Dr. Fábio Manoel França Lobato
(Orientador - Universidade Federal do Oeste do Pará)



Prof. M.Sc. Antônio Fernando
Lacerda Jacob Jr.
(Coorientador - Universidade Estadual do Maranhão)



Prof. Dr. Carlos Donato Marcílio
(Universidade Federal do Tocantins)



Prof. Dr. Quirino André Carneiro
Cortes
(Instituto Federal de Educação, Ciência e Tecnologia do Maranhão)

Assi meno più, Accendete a Roma.

AGRADECIMENTOS

Agradeço a Deus por todas as bênçãos durante esta jornada, pela saúde, família e por toda a força para superar os momentos mais difíceis.

Agradeço aos meus pais, Joiceia e José, por tudo que sempre me deram todo apoio, incentivo e estímulo para a conclusão desta jornada. E também, agradeço às minhas irmãs, Kênia, Karine e Isabella, por todo apoio e incentivo.

Agradeço à minha amada esposa, Betânia, por todo apoio, paciência e compreensão em todos os momentos difíceis desta caminhada.

Agradeço ao meu orientador Prof. Dr. Válio Leão, pelo comprometimento e oportunidades dadas desde a graduação, por todo conhecimento repassado, paciência, apoio, correções e incentivo para a conclusão deste projeto.

Agradeço ao meu coorientador Prof. Msc. Antônio Fernando Lucatela Jacó Jr., pela recepção na universidade, contribuições e apoio a todo trabalho.

Agradeço ao SOCH (Social CRM Research Center), por todo suporte e auxílio na execução deste projeto.

Agradeço à UEMA, a todos os docentes, direção e administração por todo o suporte.

Agradeço aos meus irmãos na fé da Igreja Batista Nova Jerusalém (Santarém), em especial ao Pr. Damião, Gregório Mateus e Jesus. Que através de suas mensagens de vida, foram fundamentais para minha formação e para compreensão do real sentido de fazer o que Deus.

RESUMO

A popularização dos plataformas de mídias sociais tem transformado a relação entre empresas e consumidores. Por meio destas plataformas, consumidores e empresas podem interagir, colaborar, criar e compartilhar conteúdos de forma simples, rápida e barata. Esta produção lateralizada de conteúdos possui grande influência na tomada de decisões de empresas por consumidores, pois conta de dois lados: dados climam as opiniões e avaliações de produtos e serviços antes de adquiri-los. Negligenciar este cenário pode impactar negativamente a operação diária de uma empresa, porém quando bem administrado, tal cenário representa uma rica fonte de conhecimentos sobre clientes, produtos e serviços. Nesse sentido, o Social CRM se apresenta como uma estratégia de negócios apoiada por processos e tecnologias que permitem a integração das mídias sociais a sistemas tradicionais de CRM. Nesse sentido, a utilização do Social CRM é essencial para a integração e utilização do conhecimento existente em todos os processos e áreas operacionais da empresa, o que na prática permite criar e manter os clientes mais satisfeitos. No entanto, a falta de extensão de conhecimentos a partir de conteúdos se estabelecer como uma atividade não trivial, os desafios de utilizar grandes volumes de dados na forma não estruturada e de múltiplos fontes distintas, são exemplos de pontos que limitam a plena utilização por empresas. Diante disto, esta dissertação apresenta um compêndio de quatro artigos que analisam pontos de melhoria em sistemas Social CRM, por meio da otimização da análise da estrutura das comunicações corporativas em publicações no Facebook e da utilização de dados de plataformas de redes sociais online para extrair os conhecimentos. Estes estudos permitem a exploração e a definição da melhor abordagem para a automação de análise de publicações de empresas, bem como a demonstração da viabilidade, possibilidades, vantagens e aplicabilidade em um cenário real de uso de dados de redes sociais para se aperfeiçoamento de estratégias de Social CRM.

Palavras-chave: Mídias Sociais, Marketing de Texto, Aperfeiçoamento de Mídias, Gestão de Relacionamento com Clientes, Social CRM.

ABSTRACT

The popularization of social media platforms has transformed the relationship between companies and consumers. Through these platforms, consumers and companies can interact, collaborate, create and share content in a simple, fast and cheap way. This interactive content production greatly influences purchasing decisions, as around two thirds of consumers check the opinions and evaluations of products and services before purchasing them. Neglecting this phenomenon can negatively impact the daily operation of a company, but when managed well, this scenario represents a rich source of knowledge about customers, products, and services. In this context, Social CRM presents itself as a business strategy supported by processes and technologies that allow social media integration into traditional CRM systems. In this sense, the use of Social CRM is essential for integrating and using the knowledge extracted in all of the company's processes and operational areas, which in practice allows for customer loyalty. However, the task of extracting knowledge from content is established as a non-trivial activity; the challenges of using large volumes of data, unstructured and from multiple sources, are examples of points that limit the full use by companies. In view of this, this dissertation presents a compilation of four articles that analyze points of improvement in Social CRM systems by optimizing the analysis of the effectiveness of business communications in publications on Facebook and the use of data from online complaints platforms to extract knowledge. These studies allowed the exploration and definition of the best approach for automating the analysis of company publications, as well as demonstrating the feasibility, applicability, and restrictions of using claims data to improve Social CRM strategies.

Key words: Social Media, Text Mining, Machine Learning, Customer Relationship Management, Social CRM.

LISTA DE ILUSTRAÇÕES

Figura 1	Visão integrada das funcionalidades do Social CRM. Figura traduzida e adaptada de Reinholz e Ali (2013)	21
Figura 2	Diagrama de funcionamento do modelo CRISP-DM (Adaptado de Wirth (2000))	26
Figura 3	Fluxo de trabalho das experiências	43
Figura 4	Matriz de Confusão relacionada ao tratamento de RDD consideradas	45
Figura 5	Os resultados fornecidos por análises automáticas e manuais por categorias	45
Figura 6	Métricas de engajamento associadas às categorias de conteúdo obtidas a partir da classificação automática	46
Figura 7	Diagrama de funcionamento do modelo CRISP-DM (Adaptado de Wirth (2000))	54
Figura 8	Relação entre tipos de reclamações	60
Figura 9	Número de reclamações por dia da semana	62
Figura 10	Comparação de tentativas das reclamações nos dois plataformas	72
Figura 11	Taxa de distribuição das reclamações no Brasil	72
Figura 12	Comparação no nível de resolução entre as plataformas	73
Figura 13	Contribuição da resolução de tipos	73
Figura 14	Main Types of complaints of the University A and University B	84
Figura 15	Topic distributions of complaints	85

LISTA DE TABELAS

Tabella 1:	Tarefas e ações operacionais do Social CRM: Tabela transformada e adaptada de Heintzel e An (2013).	24
Tabella 2:	Description of the data extracted.	41
Tabella 3:	Categorias de PRS adotadas de Garibaldo, Flattin e Bertoldi (2010).	42
Tabella 4:	Categorias de ECEI descritas por Garibaldo, Flattin e Bertoldi (2010).	42
Tabella 5:	Lista de algoritmos e sua melhor parametrização considerando a presença de classificação.	43
Tabella 6:	Relação do PRS com o ECU em posts do Facebook. Nota: AN = Anomalia e P = Propriedade.	44
Tabella 7:	Descrição das redes coletadas.	55
Tabella 8:	Distribuição dos dados pelas regiões do país.	57
Tabella 9:	Números de tópicos encontrados e resolvidos.	58
Tabella 10:	Modelagem de tópicos para análise de sentimentos polarizada.	59
Tabella 11:	Tópicos por região das empresas coletadas.	61
Tabella 12:	Projeções Populacionais (IBGE, 2019a), PIB (IBGE - Instituto Brasileiro de Geografia e Estatística, 2020), Índice Anso (ANVISA, 2020) e de reclamações por Região.	62
Tabella 13:	Classificação dos graus de legitimidade.	70
Tabella 14:	Número de reclamações por empresa em cada plataforma.	72
Tabella 15:	Métricas usadas para modelagem de tópicos nas plataformas analisadas.	74
Tabella 16:	Tópicos modelados a partir das reclamações que são Comuns por e Reclame Aqui.	75
Tabella 17:	Summary information about the universe.	81

LISTA DE ABREVIATURAS E SIGLAS

ANATEL	<i>Agência Nacional de Telecomunicações</i>
API	<i>Application Programming Interface</i>
ASL	<i>Average Sentence Length</i>
ASW	<i>Average number of Syllables per Word</i>
BERE	<i>Directional Encoder Representations from Transformers</i>
BRASNAM	<i>Brazilian Workshop on Social Network Analysis and Mining</i>
CF	<i>Feedback do construtor</i>
CGR	<i>Conteúdo Gerado por Empresa</i>
CGOV	<i>Governança.gov</i>
CNN	<i>Convolutional Neural Network</i>
CP	<i>Emissão do produto concreto</i>
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
CRM	<i>Customer Relationship Management</i>
CSV	<i>Comma-Separated Values</i>
DAAD	<i>Deutscher Akademischer Austauschdienst</i>
EDC	<i>Engajamento em Conteúdo Digital</i>
eWOM	<i>Electronic Word-of-Mouth</i>
FAPEMA	<i>Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão</i>
FRBS	<i>Research Reliability Bias Score</i>
HEIs	<i>Higher Education Institutions</i>
HTC	<i>Hierarchical Text Classification</i>
IBGE	<i>Instituto Brasileiro de Geografia e Estatística</i>
IWRM	<i>International Workshop on Integrated Social CRM</i>

IES	<i>Instituições de Ensino Superior</i>
IT	<i>Informação e treino/trinamento</i>
INPI	<i>Instituto Nacional da Propriedade Industrial</i>
KNN	<i>K-Nearest Neighbors</i>
LDA	<i>Latent Dirichlet Allocation</i>
LSA	<i>Latent Semantic Analysis</i>
MINIST	<i>Ministry of Justice and Public Security</i>
NMF	<i>Non-Negative Matrix Factorization</i>
NP	<i>Anúncio de novo produto</i>
OB	<i>Marca da organização</i>
PIB	<i>Produto Interno Bruto</i>
PMI	<i>Positive Mutual Information</i>
PROCE	<i>Projeto Reserua da Cultura, Comunidade e Extensão</i>
PRS	<i>Publicidade em Redes Sociais</i>
RA	<i>Real-time A/B</i>
RF	<i>Random Forest</i>
RSIT	<i>Revista Brasileira de Sistemas e Tecnologias de Informação</i>
SA	<i>Vendas</i>
SCH	<i>Social CRM Research Center</i>
Social CRM	<i>Social Customer Relationship Management</i>
SVM	<i>Support Vector Machine</i>
SW	<i>Sucesso e satisfação</i>
TF-IDF	<i>Term Frequency Inverse Document Frequency</i>
UFMA	<i>Universidade Estadual do Maranhão</i>
UFMG	<i>Universidade Federal de Minas Gerais</i>
UFPA	<i>Universidade Federal do Pará</i>

UFPA *Universidade Federal do Pará*

UGC *User-Generated Content*

SUMÁRIO

1	Introdução	21
1.1	Contextualização e desafio	21
1.2	Objetivos	26
1.2.1	Objetivo Principal	26
1.2.2	Objetivos Específicos	26
1.3	Metodologia	27
1.3.1	Etapa 1: Entendimento do negócio	27
1.3.2	Etapa 2: Entendimento dos dados	28
1.3.3	Etapa 3: Preparação dos dados	28
1.3.4	Etapa 4: Modelagem	29
1.3.5	Etapa 5: Avaliação	29
1.3.6	Etapa 6: Entrega	29
1.4	Contributo da execução do projeto	29
1.5	Organização do trabalho	31
2	Apresentação dos trabalhos	31
2.1	Trabalho "Gerenciamento de publicidade nas plataformas das redes sociais de acordo com categorias de conteúdo" - <i>Sodetyne (2019)</i>	31
2.2	Trabalho "Análise do setor de telecomunicação brasileiro: Uma visão sobre o futuro" - <i>RISEI (2021)</i>	32
2.3	Trabalho "Análise comparativa das principais plataformas de reclamações online: implicações para análise de mídia social em negócios" - <i>ILLASNAM (2020)</i>	33
2.4	Trabalho "Gaining Insights on Student Satisfaction by Applying Social CRM Techniques for Higher Education Institutions" - <i>ICRM (2021)</i>	34
2.5	Considerações	35
3	Artigo - Gerenciamento de publicidade nas plataformas das redes sociais de acordo com categorias de conteúdo	37
3.1	Introdução	38
3.2	Metodologia	41
3.2.1	Descrição do conjunto de dados	41
3.2.2	Categorias de PDS e níveis de LSA	41
3.2.3	Estrutura Experimental	41
3.3	Resultados e Discussão	44
3.4	Conclusão	47

3.5	Agradecimentos	48
4	Artigo – Análise do setor de telecomunicação brasileiro: Uma visão sobre	
	Reclamações	49
4.1	Introdução	51
4.2	Trabalhos Relacionados	62
4.3	Metodologia	64
4.3.1	Entendimento do Negócio	64
4.3.2	Entendimento dos dados	65
4.3.3	Pré-processamento dos dados	66
4.3.4	Modelagem	66
4.4	Resultados	67
4.4.1	Extração dos dados	67
4.4.2	Modelagem de Tópicos	68
4.4.2.1	Passagem Normal	68
4.4.2.2	Análise inicial	69
4.4.2.3	Distribuição "Geo-Temporal"	69
4.5	Conclusões Finais	69
5	Artigo – Análise comparativa das principais plataformas de reclamações online: implicações para análise de mídia social em negócios	65
5.1	Introdução	67
5.2	Trabalhos Relacionados	68
5.3	Materiais e Métodos	69
5.3.1	Cadência de Dados e Pré-Processamento	69
5.3.2	Extração de Características Textuais	70
5.3.3	Classificação e Modelagem de Tópicos	70
5.4	Resultados	71
5.5	Conclusões Finais	76
5.6	Agradecimentos	76
6	Artigo – Gaining Insights on Student Satisfaction by applying Social CRM techniques for Higher Education Institutions	77
6.1	Introduction	78
6.2	CRM and Social CRM in higher education	79
6.2.1	CRM affects service quality and student satisfaction in HEIs	79
6.2.2	New potentials for understanding customer satisfaction and managing the service quality seen from Social CRM	80
6.3	Improving the understanding of negative service experiences in HEIs with analytical Social CRM techniques	81

6.3.1	Complaint and satisfaction analysis by external social media	80
6.3.2	Process design	82
6.3.3	Potential data sources	82
6.3.4	Potential methods for analysis	83
6.4	Demonstration	83
6.5	Conclusion and implications	85
7	Considerações Finais	87
7.1	Trabalhos futuros	88
7.2	Dificuldades encontradas	89
	Referências	91

APÊNDICES 104

APÊNDICE A	Artigo publicado no <i>Computer on the Beach</i> (2020) - <i>Frameworks para Análise de Mídias Sociais: Um levantamento sistemático</i>	107
APÊNDICE B	Artigo publicado no <i>ICRM</i> (2021) - <i>Social CRM as a business strategy: developing dynamic capabilities of Micro and Small Enterprises</i>	115

1 INTRODUÇÃO

A popularização e a difusão das plataformas de mídias sociais têm alterado e influenciado as ações e hábitos de indivíduos e empresas (BELLÓ-ORGAZ, JUNG, CAMACHO, 2016; SHIAU, DWIVEDI, LAI, 2018; BERTHOIN *et al.*, 2012; KAPLAN, HAENSLER, 2010). Esse cenário tem transformado a forma como as comunidades e empresas se relacionam, uma vez que as comunidades passaram a compartilhar opiniões nas mídias sociais sobre os produtos e serviços. Essa relação entre comunidades e empresas é o objeto principal deste estudo, e neste capítulo serão apresentadas temas importantes que influenciam a execução, tais como a contextualização, desafios, objetivos, contexto de execução e, por fim, a organização deste manuscrito.

1.1 CONTEXTUALIZAÇÃO E DESAFIOS

Em paralelo à expansão de acesso à internet ao mundo, as mídias sociais têm ganhado cada vez mais destaque no cenário global por meio de um crescimento significativo e constante nos últimos anos (BELLÓ-ORGAZ, JUNG, CAMACHO, 2016; SHIAU, DWIVEDI, LAI, 2018). Nesse cenário de crescimento acelerado, diversas meios de interação entre as pessoas surgiram e se popularizaram, tais como blogs, vlogs, serviços de microblog (e.g. 'Twitter'), sites de redes sociais (e.g. Facebook¹, LinkedIn²), entre outros tipos. É comum que diferentes plataformas tenham em estratégias distintas de interação entre os usuários; porém, em sua essência, a interação ocorre por meio do envio e compartilhamento de conteúdos textuais e/ou visuais (CARR, HAYES, 2015). Dentre as diversas finalidades e funcionalidades que as plataformas de mídias sociais oferecem, encontra-se também a finalidade principal de cada usuário na utilização dessas mídias, baseado em Whiting e Williams (2013) cada um dos tipos é definido a seguir:

- **Interação social** — Utilizar as mídias sociais para se comunicar e interagir com outras, através da criação de posturas com os outros usuários;
- **Busca de informações** — Utilizar as mídias sociais para se manter informado sobre os diversos temas;
- **"Passar o tempo"** — Utilizar as mídias sociais para ocupar o tempo livre e aliviar o tédio;

¹ <https://twitter.com/>

² <https://facebook.com/>

³ <https://www.linkedin.com/>

- **Entretenimento** – Utilizar as mídias sociais para se divertir e se satisfazer emocionalmente;
- **Relaxar** – Utilizar as mídias sociais para aliviar o estresse do dia a dia;
- **Compartilhar informações** – Utilizar as mídias sociais como um facilitador de comunicação;
- **Por conveniência** – Utilizar as mídias sociais por considerá-las adequadas e de fácil utilização.

Em todos os diversos propósitos pessoais de utilização das mídias sociais, o conteúdo gerado por usuários, do inglês – *User Generated content* (UGC) está presente, seja no e-commerce ou na educação (FAABO; HELMS; SPIRITT, 2011). UGC pode ser definido como qualquer forma de conteúdo criado, divulgado e consumido por usuários (KIM; JOHNSON, 2016). Considerando a perspectiva de análise, alguns autores apresentam e utilizam dois subtipos de UGC: os conteúdos gerados por consumidores e os conteúdos gerados por empresas (DUNN; HARNESSE, 2019).

Os conteúdos gerados por consumidores em ambientes virtuais recebem a classificação de boca a boca virtual, do inglês, *Electronic Word of Mouth* (eWOM) e são definidos como opiniões, reações e avaliações a respeito de marcas, empresas e serviços (SCHMALL; WILKE; ROSSMANN, 2017). Na literatura, não existe um termo específico para definir o conteúdo gerado pela empresa em ambientes virtuais. Neste trabalho, tal protagonismo será denominado Conteúdos Gerados por Empresas (CGE), e são definidos como conteúdos gerados por empresas através da divulgação de produtos, serviços, respostas a clientes e divulgação de informações relacionadas a companhia (CAVALANGE; FLATDES; BHIYANI, 2019).

Para as comunidades em geral, os conteúdos de eWOM e CGE podem representar uma rica fonte de informações para tomada de decisões de compra, já para empresas, o eWOM e as interações em CGE podem representar uma fonte de dados para análise e melhoria continuamente sobre os seus consumidores e produtos (WANG; YU, 2013; LOHAPIT et al., 2017). Assim, quando bem utilizadas pelos consumidores, as mídias sociais auxiliam na tomada de decisões de compra e incrementam a possibilidade de satisfação com produtos e serviços adquiridos (WANG; YU, 2013). Já por parte das empresas, essas plataformas podem gerar vantagens competitivas por meio de melhoria de produtos, serviços e processos internos (KUBINA; LENOIR, 2015; CONSTANTINOU; HOULISUT(PINSKY, 2016).

Diante disso, a gestão de relacionamento com clientes usando redes sociais, conhecida pelo seu acrônimo em inglês, o *Social Customer Relationship Management* (Social CRM) se destaca como uma ferramenta para a integração das mídias sociais aos processos

operacionais das empresas. O Social CRM é definido como uma estratégia de negócios apoiada por processos e tecnologias que permitem às empresas integrarem dados sociais às suas estratégias, processos e sistemas de CRM (ORRINCA-BONELA; CHALMEIA, 2016; WITTEBER; REINHOLD; ALT, 2017; REINHOLD; ALT, 2013). Além das tabelas tradicionais nos CRMs tradicionais, no Social CRM há a integração e a utilização de dados de mídias sociais em cada uma destas tabelas e funcionalidades com o objetivo obter informações relevantes sobre produtos, clientes e tendências de mercado de modo a melhorar todos os processos relacionados (ALT; REINHOLD, 2012; REINHOLD; ALT, 2013; PAASE; HELMS; SPIRIT, 2011). A Figura 1 apresenta a integração entre os sistemas tradicionais de gestão de relacionamento com os clientes e as mídias sociais.

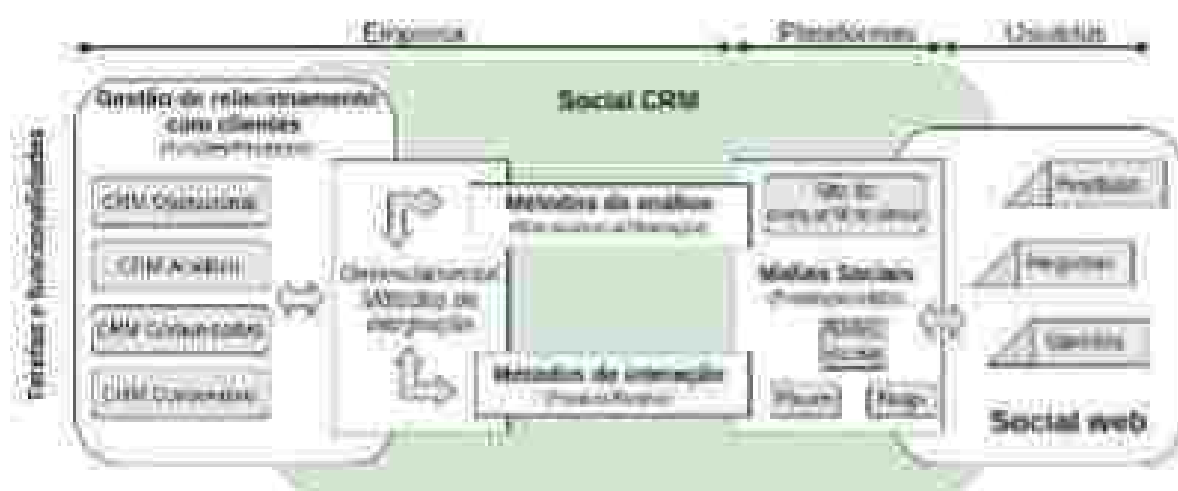


Figura 1 – Visão integrada das funcionalidades do Social CRM. Figura adaptada e adaptada de Reinhold e Alt (2013)

A integração de relacionamento de dados provenientes de mídias sociais é um processo muito importante no exercício do Social CRM (MARDET; ZIMMERMAN; PUCHNIK, 2018). Para que a integração dos dados ocorra de forma eficiente, as várias tabelas e áreas operacionais que compõem o Social CRM devem ser integradas (ALT; REINHOLD, 2012; REINHOLD; ALT, 2013). Destaca-se que processos inerentes à gestão de relacionamento com clientes poderão utilizar os resultados obtidos a partir de pesquisas e análises de relacionamento e análise do comportamento dos clientes, tal como apresentados nos trabalhos de (ORRINCA-BONELA; CHALMEIA, 2016; LOHATO *et al.*, 2017; ROSENBERGER, 2015; MARSHALL; MURK; SMOCKLEY, 2015; WITTEBER; REINHOLD; ALT, 2017). Estas tabelas e áreas listadas na Figura 1 encontram-se descritas na Tabela 1.

Tarefas/ áreas operacionais	Descrição e objetivos principais
Mídia Social	São serviços de mídia social, como fóruns, blogs, grupos para compartilhamento de informações e criação colaborativa.
Análises	Técnicas analíticas para análise e mineração, tal como filtragem, pesquisa, agregação, classificação e perfis.
Gerenciamento	Funcionalidades de gerenciamento, tal como moderação, gerenciamento de perfis, gerenciamento de reputação, integração de dados, gerenciamento de privacidade e coordenação.
CRM	Processo de CRM e integração de dados e inteligência com funções de CRM, como lead, contato, campanhas ou gerenciamento de serviço.
Interação	Técnicas de interação, tal como entrega de conteúdo, desenvolvimento de diálogo, publicação, disseminação, recomendação e coordenação de interação.

Tabela 1 – Tarefas e áreas operacionais do Social CRM. Tabela traduzida e adaptada de Reinhold e Ali (2013).

Análise de diferentes iniciativas e uso de ferramentas de Social CRM na literatura relacionada, foi insight importante quanto as formas mais adequadas de utilização das mídias sociais exposta nos estudos de (ALMEIDA, LOUNTO, CIRQUEIRA, 2017; IAMAN, MENON, 2016; CHOPRA, STAVROS, DOSSLE, 2018; OLIVEIRA, CARAS, 2018; SASIE, HERNILSEN, OLGURAN, 2019; ZHANG, YANG, 2019). Por exemplo, percebe-se que uma implementação bem-sucedida tem o potencial de contribuir com o desenvolvimento de ações preventivas para mitigar falhas e ganhar mais competitividade por meio de clientes mais satisfeitos e de fácil acesso (ORENJA, RUIA, CHALABEA, 2016).

Neste contexto, o Brasil se apresenta com um enorme potencial para as empresas utilizarem o Social CRM. Há no país cerca de 210 milhões de habitantes, sendo a região Sudeste com aproximadamente 12%, a região Sul com 14%, a região Centro-Oeste com 16%, a região Norte com 16% e a região Nordeste com 27% da população brasileira (IBGE, 2019a). De acordo a telecomunicações, de acordo com dados da Agência Nacional de Telecomunicações (ANATEL) (2020), no mês de abril de 2020 o Brasil atingiu o número de 33 milhões de conexões à banda larga fixa, o que se traduz em uma densidade de acesso de 47,4 a cada 100 domicílios, e conta com cerca de 225,6 milhões de acessos à telefonia móvel, com uma densidade de acesso de 92,9 a cada 100 habitantes. Além disso, em 2019 aproximadamente 143,5 milhões de pessoas têm acesso à internet de alguma maneira, seja por smartphone (98,6%), computadores (8,2%), telefones (21,9%) ou tablets (10,9%) (IBGE, 2019b).

Em 2020, o Brasil cresceu em 6,1% o número de usuários de internet, os quais permaneceram conectados diariamente por cerca de 10 horas e 8 minutos, divididos em 5 horas e 17 minutos na internet móvel e 4 horas e 13 minutos na internet do computador. Além disso, as plataformas de mídias sociais têm bastante relevância no país, com 70,3% da população presente em alguma plataforma neste período, representando um crescimento de 7,1% em relação ao ano anterior (Simon Kemp, 2021). Neste mesmo ano, um brasileiro gastou diariamente cerca 3 horas e 42 minutos nas suas plataformas (a terceira posição global (KEMP, 2021)). O número da população do acesso à Internet também é visto no crescimento significativo do comércio eletrônico no Brasil, que somente em 2020 cresceu 41% em relação ao ano anterior e chegou cerca de 87 bilhões de reais. Ainda neste ano, uma segunda crescerá cerca de 28% no número de clientes e passam a ter aproximadamente 79,7 milhões, os quais realizaram mais de 194 milhões de pedidos (EUF, 2021).

Além das vantagens e oportunidades que o crescimento da internet e das mídias sociais proporciona para diversas segmentos do mercado, novas desafios são impostos e potencializados. Este cenário tem transformado a maneira como os consumidores tomam as decisões de compra, pois cerca de dois terços deles aderiram às avaliações de produtos, serviços e marcas antes de decidir-se adquiri-los (AHMAD, LAHACHE, 2015; CONSTANTINIDES, HOLLESCHOVSKY, 2016). Desta forma, a percepção de um possível consumidor tende a seguir a percepção majoritária de outros consumidores nos diversos canais disponíveis para a divulgação e acesso a conteúdo de eWOM, tais como: Twitter⁴, Facebook⁵, TripAdvisor⁶, Booking.com⁷, Review Agent⁸, Comunidade.gol⁹, entre outras (FRONILLAS, VELLES, 2012; CONSTANTINIDES, HOLLESCHOVSKY, 2016).

Neste contexto, a tarefa de identificação de sentimentos a partir do conteúdo de eWOM se estabelece como uma atividade não trivial, pois existem desafios que limitam sua plena utilização por empresas, com destaque para os seguintes desafios:

- **Múltiplas fontes** - Um cidadão brasileiro tem cerca de 9 perfis em diferentes plataformas de mídias sociais (GLOBALWIRENEX, 2020). Desta forma, há a possibilidade de que o conteúdo de eWOM seja publicado em qualquer uma destas plataformas, o que requer a fusão e o tratamento de múltiplas fontes de dados distintas para a realização do processo de análise e extração de sentimentos (FARSIK, CHUA, 2017; WANG et al., 2019).
- **Dados não-estruturados** - Estima-se que cerca de 90% dos dados na internet são não-estruturados e, devido à natureza das mídias sociais, os dados são essencialmente

⁴ <https://twitter.com/>

⁵ <https://pt-br.facebook.com/>

⁶ <https://www.tripadvisor.com.br/>

⁷ <https://www.booking.com/>

⁸ <https://www.reviewagent.com.br/>

⁹ <https://comunidade.gol.br/>

tão estruturados em diferentes tipos, tal como textos, imagens, vídeos etc (KUMAR, DARIAS; HODGA, 2018; AGHASIAN, GARG, MONTGOMERY, 2019). Esta característica das mídias sociais torna os conjuntos de dados complexos e de difícil análise (CHEN, B.L; CHIANG, C. S. Sney, 2018; AGHASIAN, GARG, MONTGOMERY, 2019).

- **Dados Esparsos e Ruidosos** - Os dados provenientes das plataformas geralmente contém lacunografias e aspectos que não são úteis para as análises, por exemplo, em conteúdos textuais do português brasileiro é comum encontrar as palavras *éééé*, *carrrrrrr*, *hahahah* que indicam risos, piadas, há também outras palavras que recebem um tratamento especial, tais como *re*, *há*, *lém* e *brg*, por exemplo (FERREIRA et al., 2018). Assim, do tempo necessário para a extração de conhecimento, estima-se que 80% é utilizado somente na preparação dos dados para as análises (PHEN, 2016).
- **Grande quantidade de Dados** - Devido à inserção das plataformas de mídias sociais no cotidiano da sociedade, a quantidade de dados gerados diariamente é gigantesca, seja em fotos, textos, vídeos etc. Por exemplo, a cada minuto de 2019, cerca de 511.280 tweets foram publicados no Twitter, 35.180 fotos foram publicadas no Instagram e 92.340 posts foram publicados no Tumblr (DORMO, 2019). Assim, a grande quantidade de dados nessas plataformas é um dos principais desafios para a extração de conhecimento.

1.2 OBJETIVOS

1.2.1 OBJETIVO PRINCIPAL

Propor e avaliar diferentes abordagens para análise de dados de mídias sociais que auxiliem na melhoria de sistemas de Social CRM.

1.2.2 OBJETIVOS ESPECÍFICOS

1. Automatizar a classificação de conteúdos de marketing virais em mídias sociais em categorias de acordo com o modelo proposto por Gonçalves, Martins e David (2018), bem como correlacionar as categorias de conteúdos com métricas/padrões de engajamento. (Exposto no Capítulo 3).
2. Analisar o potencial uso de dados de reclamações para melhorar sistemas de Social CRM. (Exposto no Capítulo 4).
3. Identificar oportunidades e desafios que plataformas de reclamação podem oferecer para aproveitamento de dados de relacionamento com o cliente. (Exposto no Capítulo 5).

4. Demonstrar a aplicabilidade da análise de relações públicas nas mídias sociais em um cenário real de uso do Social CRM. Exposto no Capítulo 4.
5. Mapear o estado da arte e o estado da prática em relação à análise de mídias sociais, para identificar as bases, métodos e ferramentas mais utilizadas pelas pesquisadores em suas análises. Exposto no Apêndice A.
6. Analisar o uso de Social CRM por Micro e pequenas empresas, a fim de identificar oportunidades de intervenção. Exposto no Apêndice B.

No Capítulo 2, serão apresentados os pontos principais dos estudos que compõe o corpo desta dissertação, associados aos objetivos específicos 1, 2, 3 e 4. Desta forma, os vínculos de publicação, problemas, justificativas, objetivos, impactos, e as contribuições teórico-científicas serão destacados. Além disso, os objetivos específicos 5 e 6 são considerados estudos marginais e, portanto, estão expostos os artigos completos nos apêndices deste trabalho.

1.3 METODOLOGIA

No desenvolvimento deste estudo, foi utilizada a metodologia de análise de dados denominada *Cross Industry Standard Process for Data Mining (CRISP-DM)* (CHINCHILLA, FERREIRA, 2006; HOLLINS, 2015; SCHAFER et al., 2019; WHITE, 2000). Esta metodologia foi escolhida devido à sua maturidade e aplicabilidade em diversos cenários de análise de dados, que se adapta bem aos diversos tipos de dados disponíveis nas mídias sociais. Além disso, devido à natureza cíclica, o processo de mineração de dados pode não ser finalizado quando uma modelagem é realizada. Se for verificado que os resultados não são satisfatórios, o processo de análise pode ser reiniciado para que sejam feitas melhorias com base nas lições aprendidas em cada etapa (WHITE, 2000).

O CRISP-DM possui um processo hierárquico através de seis etapas essenciais que formam um ciclo de análise (CHINCHILLA, FERREIRA, 2006; HOLLINS, 2015; SCHAFER et al., 2019; WHITE, 2000). Na Figura 2 contém a organização de cada etapa no fluxo de execução das análises, iniciando na fase de “Entendimento do Negócio”, seguida por “Entendimento dos dados”, “Preparação dos dados”, “Modelagem”, “Avaliação” (após a avaliação também é possível o retorno para a fase inicial) e “Implantação”. Nas próximas subseções, cada uma das etapas que compõem a metodologia será descrita.

1.3.1 ETAPA DE ENTENDIMENTO DO NEGÓCIO

Nesta etapa inicial é a fase central de toda a execução das análises, pois trata os aspectos que envolvem o projeto sendo analisado e compreendidos para que seja criada uma boa definição do problema, dos objetivos e dos requisitos da solução que deve ser

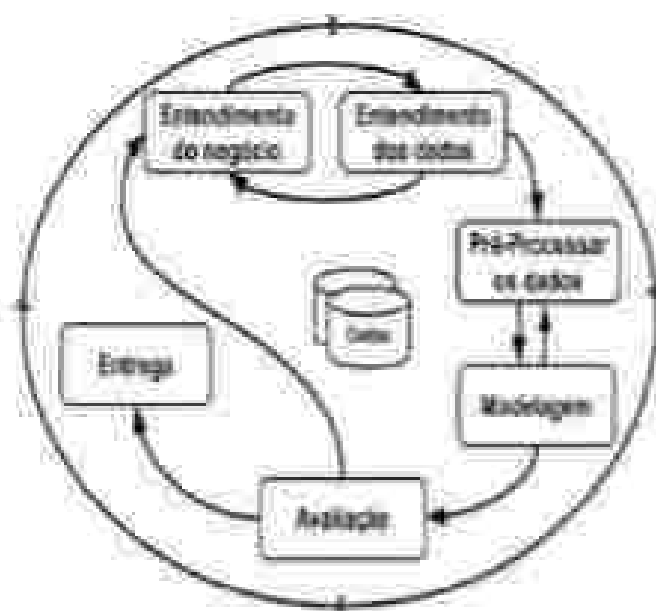


Figura 2 Diagrama de funcionamento do modelo CRISP-DM (Adaptada de [WHITE, 2006]).

implementada. Para que seja possível definir, com base no problema, objetivos e requisitos, um plano preliminar para a implementação da análise de dados [WHITE, 2006; SCHAFER et al., 2009].

1.1.2 ETAPA 2: ENTENDIMENTO DOS DADOS

Esta etapa consiste em a coleta e descrição inicial dos dados necessários à realização das análises propostas. A partir da conclusão da coleta, será realizada a exploração e verificação da qualidade dos dados coletados para familiarizar-se com esse conjunto de dados e identificar problemas de qualidade que inviabilizem a execução das análises. Além disso, a execução destas tarefas podem possibilitar a descoberta de primeiras insights sobre os dados e novas possibilidades de análise [WHITE, 2006; SCHAFER et al., 2009; CHINCHILLA, FERRERA, 2016].

Segundo White (2006), existe uma ligação estreita entre o “Entendimento do negócio” e o “Entendimento dos dados”, pois a formulação do problema e o plano de execução do projeto requerem uma compreensão básica dos dados disponíveis.

1.1.3 ETAPA 3: PREPARAÇÃO DOS DADOS

Os dados serão preparados para análise de acordo com as requisitos dos pacotes de entrada dos algoritmos de aprendizagem de máquina que serão utilizados nas análises. Dependendo do plano de execução do projeto, nesta fase, os dados brutos podem ser pré-processados para remover ruídos ou informações irrelevantes, diferentes conjuntos de dados podem ser fundidos e entre outras tarefas [WHITE, 2006; ALLAHYARI et al., 2017;

(NGUYEN et al., 2016; LI et al., 2019; CHANEY; BLET, 2012; CHEN et al., 2019).

1.3.4 ETAPA 4: MODELAGEM

Com uma primeira versão dos dados já pré-processados e preparados, nesta etapa, o enfoque será no desenvolvimento de modelos descritivos e preditivos de acordo com os objetivos definidos (HOLLANS, 2015). O processo de modelagem é altamente iterativo, dada uma vantagem de análise definida, a seleção e avaliação de algoritmos são realizadas e, à medida que os primeiros resultados são gerados, também é possível avaliar e ajustar os conjuntos de dados produzidos nas etapas anteriores (WURTEL, 2010).

1.3.5 ETAPA 5: AVALIAÇÃO

Nesta etapa, o modelo e os resultados serão avaliados para verificar sua qualidade e garantir que atende de forma adequada e completa aos objetivos definidos (WURTEL, 2010). Este processo permite identificar a qualidade e a eficiência do modelo através da utilização de métricas mais adequadas e de acordo com as especificidades de cada modelo. Caso seja verificado que o modelo e os resultados não atendam aos objetivos definidos, a firma reformula para o estágio de “Entendimento da tarefa” para reiniciar o processo, e caso contrário o processo segue para a “Entrega” (HOLLANS, 2015).

1.3.6 ETAPA 6: ENTREGA

Em todas as cases de construção de modelos de análise de dados, é importante entender o que deve ser entregue no final do processo. Pelo segundo Wirtz (2009), normalmente os resultados gerados precisam ser organizado e apresentado de uma forma, o dependendo dos requisitos, poderá ser tão simples quanto gerar um relatório ou tão complexo quanto implementar um processo de mineração de dados repetitivo¹⁶.

1.4 CONTEXTO DE EXECUÇÃO DO PROJETO

Este projeto é executado no contexto do grupo de estudos e pesquisa em computação aplicada¹⁷ da Universidade Federal do Oeste do Pará - UFOPA, o qual conta com a cooperação da Universidade Federal do Pará (UFPA), Universidade Estadual do Maranhão (UEMA), e do Instituto de Sistemas de Informação da Universidade de Leipzig, Alemanha.

Este estudo também foi executado em parceria com o Social CRM Research Center (SCRC)¹⁸. O SCRC é um instituto de pesquisa com fins lucrativos sediado em

¹⁶ Destacamos: é importante destacar que os dados implementados nos processos de análise são um arquivo importante, pois fazem parte de um programa de computador e são armazenados no HD. Por esse motivo, eles não serão entregues a partir a este trabalho.

¹⁷ <http://dep.ufopa.br/dep/qualisgrupos/112561>

¹⁸ <http://www.scrc.de/>

Leipzig na Alemanha. Não são relacionados tópicos de CRM com o enfoque na integração com Social CRM. Além disso, conta por parceria com algumas universidades brasileiras com o objetivo de prover instrumentos valiosos para profissionais e estudantes sobre negócios e o Social CRM.

1.5 ORGANIZAÇÃO DO TRABALHO

No Capítulo 2 serão apresentados os pontos principais de cada um dos trabalhos que compõem este compêndio, destacando os motivos de publicação, problemas e justificativas de cada trabalho, objetivos, impactos e contribuições dentro respectivas. Os Capítulos 3, Capítulo 4, Capítulo 5 e Capítulo 6 contêm os manuscritos completos dos artigos que foram publicados em eventos e periódicos científicos. E por fim, no Capítulo 7 estão as conclusões e considerações finais.

2 APRESENTAÇÃO DOS TRABALHOS

Neste capítulo os quatro trabalhos que compõe o corpo desta dissertação são apresentados.

2.1 TRABALHO "GERENCIAMENTO DE PUBLICIDADE NAS PLATAFORMAS DAS REDES SOCIAIS DE ACORDO COM CATEGÓRIAS DE CONTEÚDO" - SODEBRAS (2019)

Autores: Gustavo Nogueira de Sousa, Isabelle da Silva Guimarães, Antônio Fernando Leonardo Lucetti Jr., Pádua Menezes-França Lobato

Veículo de publicação: *Publicado em Revista Sodebras, Volume 14 (2019)*

Texto completo: O manuscrito publicado deste trabalho está no capítulo 3 desta dissertação.

Problema e Justificativa: É sabido que a análise de conteúdo de eWOM tem o potencial de gerar importantes ganhos e vantagens competitivas. Saber a efetividade da comunicação empresarial junto aos consumidores é muito importante, por duas formas, a primeira medir o impacto das ações implementadas com base na análise de conteúdo. No entanto, a quantidade de dados gerados pelas plataformas torna o processo de avaliação manual dos resultados das publicações de negócios muito caro e ineficiente. Diante disto, este trabalho tem como objeto de estudo o melhoramento da utilização de mídias sociais como meio para o relacionamento entre empresas e consumidores, em que avalia formas de otimização do processo de avaliação da efetividade, em termos de engajamento, de publicações empresariais.

Objetivo: O objetivo principal deste trabalho é automatizar o processo de mensuração da efetividade de publicações empresariais, por meio da classificação e correlação automática das categorias de expressão com métricas/padrões de engajamento, de acordo com o modelo proposto por Cavallaro, Platten e Dwyer (2015). Além disso, alguns objetivos secundários também estão implícitos no trabalho, tais como:

- Avaliar o desempenho de algoritmos de aprendizado de máquina em tarefas de classificação de conteúdos de mídias sociais;
- Eliminar o viés humano na classificação dos conteúdos das publicações;

¹ <http://sodebras.com.br/>

- Prover um guia para a classificação e avaliação da efetividade de grandes volumes de dados;
- Tornar o processo de avaliação da efetividade de publicações menos custoso para empresas.

Impacto e contribuições técnicas científicas Este trabalho propõe e valida uma abordagem para as avaliações de publicações de empresas via mídias sociais. Os resultados são pertinentes, pois elimina o viés da classificação e avaliação manual dessas publicações, tornando o processo mais eficiente e menos custoso, o que permite a tomada de decisões sobre a forma que está sendo construída o relacionamento com os clientes. Além disso, este trabalho foi necessário a criação manual de uma base de dados de treinamento com publicações empastilhadas categorizadas de acordo com as categorias de conteúdos proposta por Cavallaro, Poles e Herold (2016). Esta base de dados foi tornada pública (SOUSA, JUNIOR, LOPES, 2021) e pode ser usada como ponto de partida para a utilização em outras análises pela comunidade.

2.2 TRABALHO “ANÁLISE DO SETOR DE TELECOMUNICAÇÃO BRASILEIRO: UMA VISÃO SOBRE RECLAMAÇÕES” - RISTI (2020)

Autores: Gustavo Nogueira de Sousa, Isabella da Silva Guimarães, João Augusto Nogueira Vianna, Otávio Heilmann, Antonio Fernando Laurente Junior Junior, Fábio Manuel Franco Lobato.

Veículo de publicação: Publicado em periódico *Revista Brasileira de Sistemas e Tecnologia da Informação (RISTI)* (2020).

Texto completo: O manuscrito publicado deste trabalho está no capítulo 3 desta dissertação.

Problema e Justificativa: Os conteúdos de eWOM representam uma rica fonte de informação para clientes, visto que cerca de dois terços dos consumidores verificam as avaliações de produtos e serviços antes de decidirem adquiri-los (AJMALI, LARSEN, 2017; CONSTANTINIDES, HILLESCHROVSKY, 2016). No entanto, pesquisas na literatura revelam que há uma grande variedade de trabalhos que analisam dados de eWOM de diversos plataformas de mídias sociais (KIM, JOHNSON, 2016; VERMEER et al., 2019; MCILROY et al., 2016; VI et al., 2016), porém são poucos os trabalhos que usam conteúdo eWOM publicado nas plataformas específicas para reclamações online. No Brasil, as plataformas de reclamações online têm bastante relevância e influência, sendo que a principal plataforma do tipo está entre as 25 sites mais acessados do país. Diante disso,

² <https://www.risti.org/>

este trabalho tem como objeto de estudo a utilização dessas plataformas como fonte de dados para a extração de conhecimentos.

Objetivos: O objetivo principal deste trabalho é o de analisar o potencial uso de dados de reclamações para melhorar os sistemas de Social CRM. Para isso, o objetivo principal está segmentado nos objetivos específicos a seguir:

- Identificar as principais temáticas presentes nas reclamações;
- Identificar como as principais tópicos estão relacionados entre si;
- Classificar aspectos específicos das reclamações;
- Analisar a distribuição das reclamações considerando dimensões pre-terceiras;
- Identificar quais as implicações práticas dos resultados das análises realizadas para os negócios.

Impacto e contribuições acadêmicas científicas: Apesar das análises realizadas, este trabalho demonstra a análise de reclamações como uma forma viável de aprimorar os sistemas de Social CRM. A abordagem utiliza técnicas de aprendizado de máquina, que permitirão identificar e mensurar o grau de associação entre as principais temáticas nas reclamações dos consumidores. Além disso, neste trabalho inicia-se a construção de uma pipeline de análise com todas as técnicas utilizadas, com objetivo de reduzir o tempo requerido para as análises. Os resultados deste trabalho foram avaliados por um especialista em desenvolvimento de negócios do Social CRM Research Center³, que atesta a viabilidade de uso para a melhoria dos relacionamentos com clientes.

2.3. TRABALHO "ANÁLISE COMPARATIVA DAS PRINCIPAIS PLATAFORMAS DE RECLAMAÇÕES ONLINE: IMPLICAÇÕES PARA ANÁLISE DE MÍDIA SOCIAL EM NEGÓCIOS" - BRASNAM (2019)

Autores: Gustavo Nogueira do Sousa, Isabella Guimarães, Antonio P. L. Jacobi Jr., Rômulo M. F. Ladeira.

Voluntade de publicação: Publicado em anais do Brazilian Workshop on Social Network Analysis and Mining (BrASNAM) (2019)⁴

Texto completo: O manuscrito publicado deste trabalho está no capítulo 5 desta dissertação.

³ <https://scm.bepko.de/>

⁴ <http://www.abn.org.br/abn2019/pt-br/brazilian-workshop-on-social-network-analysis-and-mining/>

Problema e Justificativa: Apesar da presença das plataformas de mídia social no cotidiano da sociedade, as comunidades possuem a sua responsabilidade pelo consumo quanto pela produção de conteúdos sobre um produto/serviço nas mídias sociais (ROZ, DUTTA, MUKHIGUJE, 2019). Devido a facilidade de propagação e de engajamento nessas plataformas, tais conteúdos têm a capacidade de influenciar a tomada de decisões de compra de milhões de clientes (CYRSTANTENIDES, HOLLESCHOWSKY, 2016). Por meio de pesquisas na literatura, percebeu-se que há uma escassez de estudos que analisem plataformas especializadas para reclamações online, portanto, este trabalho tem por objeto de estudo um comparativo entre duas das principais plataformas de reclamações online do Brasil.

Objetivo: O objetivo principal deste trabalho é identificar oportunidades e desafios que plataformas de reclamação possuem para aperfeiçoamento de perfil de relacionamento com o cliente. Para isso, alguns objetivos específicos também foram definidos, tal como:

- Determinar de forma objetiva a quantidade mais adequada de cópias da modelagem;
- Analisar através da modelagem de cópias as diferenças entre as conjunções de reclamação de cada empresa nas plataformas;
- Analisar as diferenças dos grupos de comunidades nas duas plataformas.

Impacto e contribuições técnicas científicas: Este trabalho apresenta as possibilidades e contribuições de análises de diferentes plataformas de reclamações online para a melhoria de atuação de Social CRM. Desta forma, apresenta para o português brasileiro a aplicação de um método automatizado para determinar de forma objetiva a melhor quantidade de cópias para modelagem, bem como as diferenças entre as duas principais plataformas e as diferenças entre os grupos de comunidades presentes nessas plataformas. Os métodos utilizados no processo de extração de informações possibilitam que melhorias e novas funcionalidades possam ser adicionadas ao pipeline de análise elaborado no artigo descrito na Seção 2.3.

2.4. TRABALHO "GAINING INSIGHTS ON STUDENT SATISFACTION BY APPLYING SOCIAL CRM TECHNIQUES FOR HIGHER EDUCATION INSTITUTIONS". ICRM (2021)

Autores: Gustavo Nogueira de Sousa, Pablo Lobato, João Vilas, Odi Reichert.

Vetendo de publicação: Publicado em anais do International Workshop on Intelligent Social CRM (ICRM 2021)

² <https://www.icrm.org/2021/track/>

Texto completo: O manuscrito publicado deste trabalho está no capítulo 6 desta dissertação.

Problema e Justificativa: Por meio de pesquisas na literatura, foi possível verificar que existem poucos estudos sobre gestão de reclamações por Instituições de Ensino Superior (IES), e também que o potencial de construção de relacionamentos a partir do percurso do aluno é apenas parcialmente examinado na literatura. Diante disso, este trabalho tem como objeto de estudo a utilização de uma plataforma de reclamações online para melhorar o percurso acadêmico dos alunos nas IES.

Objetivos: O objetivo principal deste trabalho é demonstrar a aplicabilidade, em um cenário real de uso do Social CRM, da análise de reclamações publicadas nas mídias sociais. Para isso, a relevância das mídias sociais é explicada e analisada como canal para tratamento de reclamações como parte do processo de melhoria da trajetória acadêmica de estudantes em IES brasileiras. Para tanto, foram definidos os seguintes objetivos:

- Explorar a utilização das mídias sociais para reclamações por alunos;
- Analisar se há uma ligação entre o número de reclamações e a gestão ativa das redes sociais;
- Identificar os principais temas das reclamações;
- Identificar a relação entre os diferentes perfis das universidades.

Impacto e contribuições técnicas científicas: Por meio das análises realizadas, este trabalho demonstra a aplicabilidade da análise de mídias sociais na melhoria da trajetória acadêmica em IES. A abordagem utilizada e usada as técnicas de aprendizado de máquina implementadas no pipeline de análise nos trabalhos apresentados nas Seções 1.2 e 2.2 e permitem o tratamento e a extração de conhecimentos das reclamações coletadas. Os resultados deste trabalho foram analisados por um especialista em desenvolvimento de negócios e por um especialista em Social CRM, ambos do Social CRM Research Center⁸, que atestaram a validade da utilização de dados de reclamações para melhorar a trajetória acadêmica.

2.5. CONSIDERAÇÕES

Neste capítulo, foram apresentadas as partes principais de cada um dos quatro capítulos que compõem a parte desta dissertação. Cada capítulo está associado a um dos objetivos específicos descritos na Seção 1.2 e abrange diferentes perspectivas para a análise de dados de mídias sociais usada para negócios. Foi utilizada como base

⁸ <http://www.socp.org.br/>

metodológica o CRISP-DM, tal como o processo descrito na Seção 1.2, o que possibilitou a execução eficiente de diferentes perspectivas e condições de análise. Estes processos de análise e os resultados obtidos mostraram-se relevantes e têm o potencial para aprimorar a utilização do Social CRM.

Desta forma, os estudos realizados nos artigos, os quais foram descritos neste capítulo, Com propósito de prover uma visão detalhada dos resultados obtidos, nos Capítulos 3, 4, 5 e 6 contém o texto completo de cada um dos artigos.

3. ARTIGO - GERENCIAMENTO DE PUBLICIDADE NAS PLATAFORMAS DAS REDES SOCIAIS DE ACORDO COM CATEGORIAS DE CONTEÚDO

RESUMO

O uso de mídias sociais está se expandindo por diferentes setores da sociedade, consequentemente, uma grande quantidade de conteúdos gerados pelos usuários é produzida todos os dias. Devido aos diferentes tipos de conteúdos, a gestão de conteúdo é essencial para a publicidade comercial nas mídias sociais. No entanto, o grande volume de dados faz com que se crie uma dificuldade para a análise dos dados que se encontram disponíveis em escala em redes sociais. Este artigo investiga o uso de técnicas de gerenciamento de mídias sociais para automatizar o processo de análise, aumentando a eficiência dos processos e a confiabilidade dos resultados. Mais especificamente, analisa-se o uso de um classificador de texto para categorizar as publicações de acordo com o seu conteúdo, como estudos de caso, anúncios de publicidade de empresas no Facebook. Os resultados mostram que o classificador obtido apresenta potencial para analisar uma quantidade significativa de conteúdos com eficiência. O classificador tem implicações práticas, pois os que permitem uma melhor análise dos conteúdos e também é capaz de influenciar as campanhas de marketing em mídias sociais.

Palavras-chave: Rede Social, Propaganda, Empacotamento de mídias, Aprendizado de Máquina, Técnicas de Decisão Baseada em Dados.

ABSTRACT

Social media usage is expanding in different sections of society; consequently, a large amount of User Generated Content is produced every day. Due to its various effects on users, content management is essential for business advertising on these platforms. However, the massive social media's data volume, increases the costs for analyzing the content effects on users. This paper examines the use of machine learning techniques to reduce the cost and effort of this kind of analysis. More specifically, an automatic document classification to identify content categories is evaluated. As a case study, we adopted some Facebook companies' posts. The results show that the machine learning classifier obtained has the potential to analyze a significant amount of content. The classifier has practical implications since it allows an extensive competitor analysis to be conducted and is also able to influence social media campaigns.

Keywords: Social Network, Advertising, User Engagement, Machine Learning, Data Driven Decision Making.

3.1 INTRODUÇÃO

O uso de mídias sociais está crescendo constantemente (BELLACHIOGAZ, JUNG; CAMACHO, 2016). Em 2019, estima-se que essas plataformas tenham cerca de 2,77 bilhões de usuários, com o Facebook como plataforma líder, com 2,3 bilhões de usuários (SHIBAU; DWIVEDI; LAI, 2018). Além disso, muitas plataformas apresentam um grande número de usuários, como o YouTube com 1,8 bilhões, o Twitter com 330 milhões e o Instagram com 1 bilhão (STATISTA, 2019). Devido à facilidade de usar, os usuários podem criar, publicar, colaborar e compartilhar conteúdo com outras pessoas por meio dessas mídias (MAIZ; ARRIANZ; JUAN, 2013). Essa facilidade também vem sendo utilizada significativamente na comunicação e interação social, impactando diretamente na relação entre as empresas e seus clientes (PRADHYUMNI, 2011; ALMEIDA; LORITO; CIRQUEIRA, 2017; LORITO et al., 2017; Nogueira de Sousa et al., 2018).

Devido à facilidade para compartilhar informações nessas plataformas, tornou-se um fenômeno chamado de Boca a Boca Virtual (electronic Word-of-Mouth - eWoM), que transforma as comunicações em alguns sites (TARHOTO, 2014). O eWoM pode ser entendido como o ato de criar e compartilhar informações sobre produtos, serviços e serviços nas mídias digitais (SCHMIDT; WILKE; HERSMANN, 2017). Isso significa que a mídia social se tornou um importante meio para compartilhar seus tipos de conteúdo (AHMAD; LAROCKHE, 2017), podendo representar uma importante fonte de informações sobre as preferências das pessoas (BENSON, 2010).

As informações derivadas do eWoM permitem uma comparação entre produtos, produtos ou serviços (HUSSAIN et al., 2016). Ao aplicar métodos de detecção de comunidades em dados de mídia social, é possível aprender insights úteis sobre algumas das dinâmicas e fenômenos que ocorrem nesses sistemas (SILVA et al., 2017). Além disso, este fenômeno está diretamente relacionado ao setor de turismo, já que um grande número de turistas seleciona seu destino, hotel, passagens e restaurantes com o auxílio de conteúdos do eWoM, tais como fotos, vídeos, análises e avaliações (HARRISMAN et al., 2017; OLIVEIRA; CARVAL, 2018). Isso pode ser descrito como Turismo Inteligente, que é definido como atendimento ao cliente de forma imersiva por meio de informações turísticas relevantes, e é caracterizado pela privacidade, gerenciamento e compartilhamento de serviços e experiências durante a jornada dos turistas (CHITZEL et al., 2015; Li et al., 2017).

As transformações e mudanças trazidas pelo uso do Turismo Inteligente podem ser potencializadas através da integração de sistemas de CRM (Customer Relationship Management) e planejamento de estratégias de atuação nas mídias sociais (CHUJIM; PALACIOS et al., 2017). Essa estratégia inclui os dados coletados na jornada do cliente e leva à transparência e confiança, tem resultados positivos com os clientes (CHUJIM; PALACIOS et al., 2017; VIECHIO et al., 2016).

Tendo em vista a importância do eWOM para o mercado, neste artigo focamos na Publicidade em Redes Sociais (PRS) no que diz respeito a: A) As formas de conteúdo que são criadas pelas marcas e disseminadas pelas mídias sociais; e B) o Engajamento em Conteúdo Digital (ECD), que pode ser definido como o estado psicológico induzido pelas interações com a identidade da marca em um ambiente digital. Sete categorias de conteúdo PRS e três níveis de ECD são definidos por [Garcia, Platten e Bretzel \(2018\)](#), que podem ser correlacionados pelas análises de marketing para determinar sua eficácia. Devido a grande quantidade de conteúdo PRS nas mídias sociais, são necessárias ações rápidas e um esforço considerável para analisar seu impacto no ECD ([Ull et al., 2017a](#)). Neste contexto, a seguinte questão de pesquisa foi definida:

- É possível a classificação automática de posts de acordo com as categorias de conteúdo de PRS apresentadas por [Garcia, Platten e Bretzel \(2018\)](#)?

Para responder a essa questão de pesquisa, primeiro devemos identificar as necessidades de aplicação de máquinas, a fim de desenvolver um classificador automático de conteúdo de PRS com base nas categorias supracitadas. O método de classificação automática de dados foi desenvolvido e testado em publicações sobre cartões no Facebook. Além disso, os pesquisadores validaram o modelo e discutiram suas implementações práticas. Por exemplo, é possível usar o classificador na automação da análise de comentários, na implementação de estratégias de marketing, avaliando o conteúdo de PRS e seu engajamento. É possível também o desenvolvimento de um sistema de suporte à decisão para prever o engajamento do modelo com base no conteúdo da postagem.

O restante deste artigo está estruturado como segue. Na Seção 2 discute-se a metodologia empregada no estudo. Os resultados são analisados e discutidos na Seção 3. Finalmente, as conclusões e sugestões de trabalhos futuros são apresentadas na Seção 4.

3.2 METODOLOGIA

Nesta seção, a questão de pesquisa é respondida por meio de uma descrição do conjunto de dados, seguida pelas categorias de PRS e níveis de ECD, e pelo estabelecimento da Estrutura Experimental.

3.2.1 DESCRIÇÃO DO CONJUNTO DE DADOS

Neste trabalho, usamos dados que foram extraídos de publicações no Facebook¹ de diversas empresas da área de turismo. Seguindo as seguintes critérios de seleção: (i) as empresas precisavam ter perfil no Facebook; e (ii) as empresas deviam estar ativas no

¹ Declaramos: Todos os dados coletados no Facebook serão utilizados exclusivamente para fins de prova de conceito. Os autores não tem qualquer interesse no uso comercial desses dados.

últimos seis meses, em outras palavras, foi realizada no respectivo período. Esta abordagem foi adaptada de [NOGUEIRA DE SOUSA et al. \(2018\)](#). O Facebook foi escolhido devido à sua popularidade, grande número de usuários e consequente relevância mercadológica. A extração de dados foi realizada por meio da API oficial do Facebook de janeiro a julho de 2018.

Os critérios de seleção foram definidos com base em [Melo, Arruda e Juan \(2016\)](#). No total, dados de 93 empresas foram coletados de uma ampla gama de empreendimentos correlatos ao turismo, entre hotéis, restaurantes, lanchês, pousadas e assim. Ao todo, houve um total de 10.925 publicações durante o período de extração, sendo um dado para cada publicação está descrito na Tabela 2.

Tabela 2 – Description of the data extracted.

Dados	Formato	Descrição
Post ID	Números	Identificação de cada publicação na rede social
Texto	String	Conteúdo textual de cada publicação
Tipos	String	Tipo de publicação - 'foto', 'vídeo', 'status' ou 'link'.
Link	URL	Post link.
Data da publicação	Data	Publication date in the social network.
Reação	Enumeração	As reações dos usuários ao post - essas reações são: 'Curtir', 'admirar', 'amar', 'surpreender', 'triste', 'raiva', 'surpresa'.
Compartilhamento	Números	Número de vezes que a publicação foi compartilhada.

3.2.2 CATEGORIAS DE PHS E NÍVEIS DE ECD

Este artigo foca na publicidade em redes sociais e seu nível de engajamento de conteúdo digital descrito por [Livillano, Hansen e Davies \(2018\)](#). A Tabela 3 mostra as categorias de conteúdo com suas respectivas descrições. Da mesma forma, Tabela 4 mostra os níveis de ECD com uma avaliação de seu grau de influência sobre sua eficácia e a descrição de cada nível.

3.2.3 ESTRUTURA EXPERIMENTAL

A estrutura experimental adotada para o estudo é descrita na Figura 3 e é composta das seguintes etapas: 1) Aquisição de dados (dados brutos); 2) Pré-processamento de dados; 3) Análise manual de uma amostra significativa dos dados; 4) Classificação do conteúdo das publicações por meio de um algoritmo de aprendizagem de máquina; 5) Medição da precisão e validação do classificador obtido no passo 4; 6) Correlação das categorias PHS com os níveis de ECD; 7) Validação dos resultados obtidos.

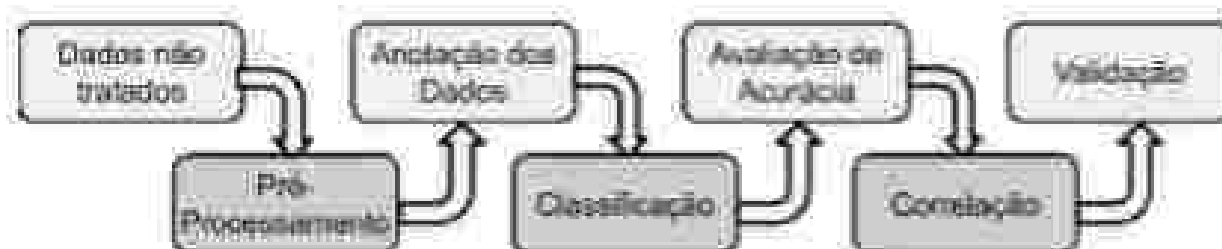


Figura 3 – Fluxo de trabalho dos experimentos.

Tabela 5 – Lista de Algoritmos e sua melhor parametrização considerando a precisão da classificação.

Algoritmos	Parâmetros	Acurácia
KNN	$k = 9$, distance = "eucl"	71 %
Gaussian Naive Bayes	Prior= None, Var_smoothing= 10^{-5}	71 %
SVM	$C = 1000$, Kernel= "sigmoid"	86 %
Multinomial Naive Bayes	Alpha= 1.0, Class_prior= None, Fit_prior= True	69 %
Random Forest	n_estimators=7, min_samples_split=9	79 %

A primeira etapa, descrita na Figura 3, é a aquisição de dados, a qual foi realizada por meio da API do Facebook. O processo de classificação automatizada envolve preparar e aplicar os dados para construir um modelo de classificação. A segunda etapa foi a aplicação de métodos de pré-processamento nos textos de cada publicação visando reduzir o ruído. Para isso, usamos o workflow descrito por [Carpentier et al. \(2017\)](#). O ruído, neste contexto, era composto por URLs, palavras irrelevantes, números, acenções, emojis e caracteres especiais.

A terceira etapa do fluxo de trabalho refere-se à anotação de dados manual de acordo com as categorias de PICS mencionadas anteriormente. A anotação foi realizada por dois avaliadores independentes utilizando-se de um sistema adaptado para tal ([CHAUVERA et al., 2017a](#)). A confiabilidade das anotações foi realizada por meio da Conferência Externa de Cohen e apenas os dados com confiabilidade entre os avaliadores foram retidos, em outras palavras, somente quando os dois avaliadores atribuíram o mesmo título a uma publicação, a publicação era incluída ao conjunto de dados de treinamento.

Para a quarta parte do processo descrito na Figura 1, alguns algoritmos de aprendizado de máquina foram usados para classificações feitas testados para responder as questões de pesquisa previamente definidas. Os algoritmos testados foram: K-Nearest Neighbors (KNN), Gaussian Naive Bayes, Support-Vector Machine (SVM), Multinomial Naive Bayes e Random Forest (RF). Esses algoritmos estão disponíveis no framework scikit-learn ([DUPUYREX et al., 2011](#)). A parametrização dos algoritmos foi realizada usando GridSearch para SVM, KNN e RF. Para os outros algoritmos os parâmetros padrão foram usados. Na Tabela 5 os algoritmos e parâmetros utilizados são apresentados.

Foram utilizadas as seguintes métricas de desempenho: *Acurácia*, *Precisão*, *Recall* e *F1-measure* (ponderado), e os dados foram reutilizados por meio de validação cruzada (10 vezes). A *F1-measure* (ponderada) foi adotada em vez de *micro / macro*, uma vez que leva em conta o desequilíbrio do título. A luz das melhores métricas de desempenho, o algoritmo SVM foi adotado (Tabela 3). A correlação das categorias de PDS com as métricas que determinam os níveis de KDD é realizada para determinar a eficácia do conteúdo. A validação foi realizada considerando um cenário real relacionado ao turismo inteligente, no qual dois profissionais que atuam como *reviewer* realizaram uma avaliação qualitativa dos resultados obtidos nas etapas anteriores.

O processo de análise consistiu em uma seleção aleatória de algumas publicações classificadas para serem usadas no processo de validação. Além disso, foi aplicado um algoritmo de modelagem de lógica para todo o conjunto de dados classificados, apresentando os resultados (por classe) para as probabilidades. Das análises se a correlação da lógica era consistente com as classes propostas. No entanto, devido a restrições no tamanho do artigo, a apresentação deste processo foi suprimida dos resultados.

3.3 RESULTADOS E DISCUSSÕES

Neste trabalho foram coletados um total de 10.925 publicações de 93 empresas situadas no setor de turismo, como hotéis, restaurantes, hotéis, parques e sites. Considerando a quantidade de publicações coletadas, foram necessárias 628 publicações para obter um intervalo de confiança de 99% com uma margem de erro de 1%. No entanto, como era possível que os resultados não correspondem ao mesmo estudo para os mesmos dados, 1.020 publicações foram excluídas para amostrar manual.

Como mencionado na Seção de Metodologia, as amostras foram analisadas usando o coeficiente Kappa de Cohen e obtiveram o valor de 0,5, o que significa que há um nível moderado de concordância entre as amostras. Esses dados inicialmente representam 9% do número total de publicações, embora apenas os dados classificados com o nome *ótimo* fossem mantidos, o que resultou em 680 publicações, ou cerca de 6 %, uma das que é esperada pelo intervalo de confiança e erro previamente definido. Todos os dados foram extraídos, anotados e processados automaticamente.

Várias etapas de pré-processamento foram realizadas no conjunto de dados de treinamento, para limpar e remover todas as informações desnecessárias, por exemplo, URLs e palavras irrelevantes. No classificador SVM, as publicações que foram analisadas manualmente foram submetidas ao treinamento do algoritmo, e os resultados obtiveram uma *Acurácia* de 80,67%, *Precisão* de 77,60%, *Recall* de 80,87% e *F1-measure* de 79,17%. As métricas de desempenho, também, podem ser observadas na Figura 1, que foi normalizada para fornecer uma melhor visualização de dados.

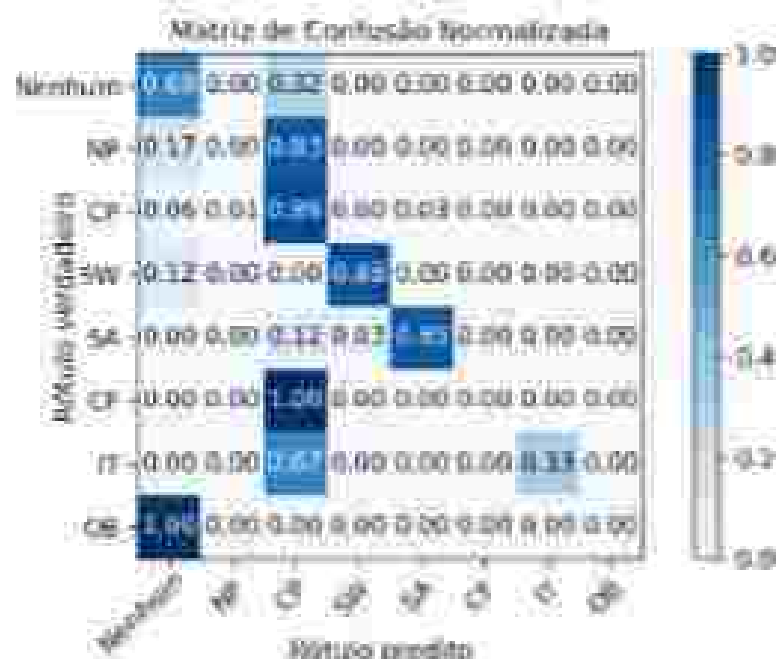


Figura 4: Matriz de Confusão relacionada às categorias de ECD consideradas.

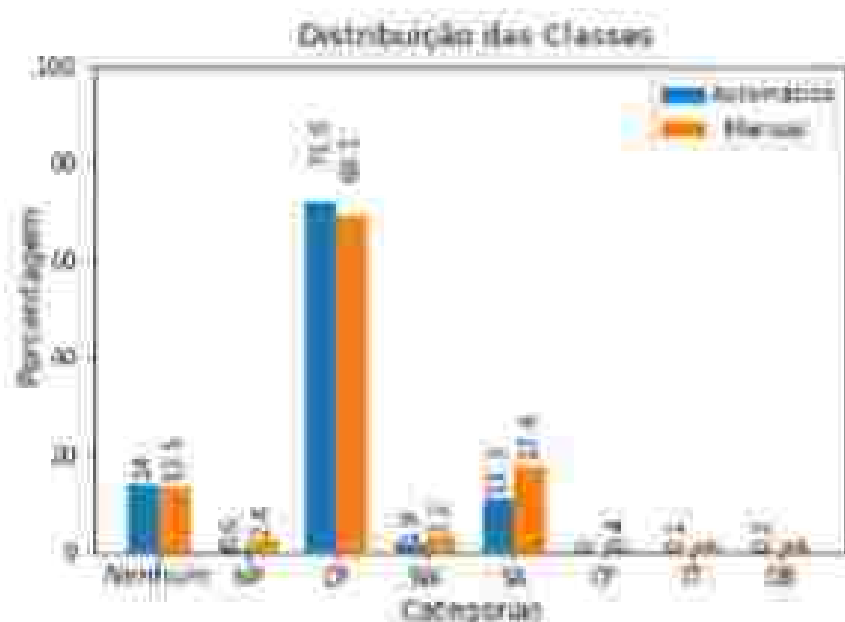


Figura 5: Os resultados gerados por abordagens automática e manual por categorias.

Como forma de prova que a classificação automática de publicações pode obter resultados semelhantes a uma classificação manual, a classificação foi aplicada ao restante do conjunto de dados (cerca de 94 % dos dados). Os resultados foram positivos, com a distribuição das categorias nos resultados observados na Figura 5 é possível observar que o uso de técnicas de aprendizado de máquina é uma maneira eficaz de reduzir o esforço da tarefa de rotulagem para medir a eficácia do PDS em ECD. Além disso, poderia analisar

grandes quantidades de dados e determinar como é a melhor maneira de analisar por meio de marketing em redes sociais (GAVILANES, FLATIES, BRETHER, 2018).

A Tabela 6 foi montada com base nos resultados da classificação automática, e uma correlação pode ser observada entre as categorias de PDS e as métricas das redes de ECD propostas por Gavilanes, Flaties e Bretter (2018). Nota-se na Figura 6, as categorias de Notícias e Comentários têm um apoio maior ao ECD de acordo com a média de curtidas, comentários e compartilhamentos em cada publicação.

Tabela 6 | Relação do PDS com o ECD em posts do Facebook. Nota: AV = Média x F = População.

	Filtragem positiva		Processamento negativo e abstrato		Apoio	
	F	Av	F	Av	F	Av
Senhora	87175	54.1	3520	2.3	248	3.4
Anúncio de novo produto	28523	146.7	947	5.5	672	1.7
Exposição do produto comercial	601910	81.5	79267	5.1	26394	3.5
Serviços e conteúdos	66405	244.4	42617	157.8	7405	157.8
Vendas	126158	100.9	3571	2.8	1742	1.3
Feedback do consumidor	311	16.5	27	1.35	1	0.2
Informações e entretenimento	1977	34	140	2.75	90	1.7
Marca da organização	7406	66.5	459	1.2	219	2

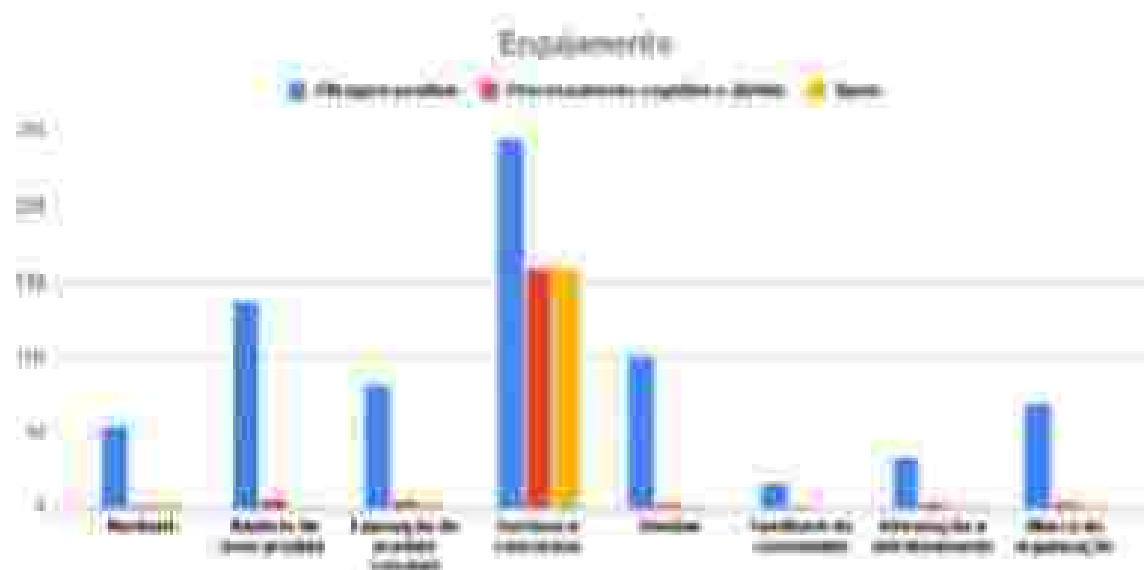


Figura 6 | Métricas de engajamento associadas às categorias de conteúdo utilizadas a partir da classificação automática.

3.4 CONCLUSÃO

Esta pesquisa investiga o uso de técnicas de aprendizado de máquina para analisar o custo e o impacto de várias estratégias de Publicidade em Redes Sociais. Para isso, adotamos as categorias de Engajamento de Conteúdo Digital propostas por [Garcimén, Flattin e Brottet \(2019\)](#) e dados do Facebook. Com base em experimentos extensivos, avaliamos vários classificadores de aprendizado de máquina, a saber, K-Nearest Neighbors (KNN), Gaussian Naive Bayes, Support Vector Machine (SVM), Multinomial Naive Bayes e Random Forest (RF). Os resultados mostram que o classificador SVM tem um excelente potencial para realizar a classificação de conteúdo, a qual permite uma posterior correlação entre categorias e recado de engajamento. Sendo capaz de analisar uma grande quantidade de conteúdo com maior eficiência e menor custo/errores. Assim, em termos proporcionais, o classificador automático pode alcançar resultados semelhantes aos obtidos pela anotação manual. No entanto, como é possível observar na Figura 4, algumas classes de conteúdo obtiveram uma classificação com erro insuportável, a saber a NP, CF, EF e OD. Isto ocorreu devido ao desbalanceamento na quantidade de dados rotulados manualmente para treinar os classificadores para estas classes, como exemplificado na Figura 5.

Os resultados dos classificadores foram avaliados e validados por profissionais. Com base nos resultados, algumas aplicações práticas foram visualizadas: i) é possível automatizar os conteúdos analisados automaticamente o conteúdo publicitário das mídias sociais; ii) o classificador pode ser incorporado em um sistema de suporte à decisão, auxiliando na implementação do engajamento do usuário e na desenvolvimento de novas estratégias de marketing; iii) o classificador também pode apoiar a análise do engajamento do usuário de acordo com as demandas atuais do mercado por gerenciamento transacional nos setores de energia e outros. Estes são serviços bem conhecidos fornecidos por empresas de marketing digital ([BARATA et al., 2019](#)), que podem ser aprimoradas pelo classificador automático.

No entanto, existem algumas limitações em nosso estudo que precisam ser abordadas em estudos futuros. Primeiro, nosso conjunto de dados inclui apenas dados de usuários e, assim, resultou em um conjunto de dados desbalanceado. Em segundo lugar, foram apenas dois analistas de dados rotulados, houve um baixo intervalo de confiança nos resultados. Em vista disso, em trabalhos futuros, pretensionamos de expandir nosso conjunto de dados para outros setores de mercado e incluir um terceiro analista para melhorar a confiabilidade da pesquisa. Além disso, estamos planejando incluir técnicas que lidam com conjuntos de dados desbalanceados ou para melhorar o desempenho da classificação.

3.5 AGRADECIMENTOS

Os autores agradecem o apoio financeiro a esta pesquisa da Universidade Federal do Oeste do Pará (UFOPA), do Serviço Alemão de Intercâmbio Acadêmico (DAAD), SORM-SP/USP e Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA).

4 ARTIGO - ANÁLISE DO SETOR DE TELECOMUNICAÇÃO BRASILEIRO: UMA VISÃO SOBRE RECLAMAÇÕES

RESUMO

Mídias digitais estão cada vez mais presentes no cotidiano de um cidadão. Este fato contribui para que o volume de conteúdo gerado por usuários continue consideravelmente. De um ponto de vista prático, as redes sociais dão origem a diferentes perspectivas e métodos para se obter informações estatísticas. Esse cenário pode auxiliar a tomada de decisões por gestores por meio da identificação de necessidades e problemas, gerando o processo de melhoria contínua de produtos e serviços. Diante disso, este trabalho propõe uma análise de reclamações postadas em uma plataforma online de reclamações, a fim de identificar pontos que causam a tomada de decisão das empresas e, consequentemente, melhorar o relacionamento com clientes. Os resultados obtidos permitem a identificação de uma cadeia de problemas relacionados. A principal contribuição deste estudo está na proposta de uma abordagem que auxilia no planejamento estratégico de empresas, levando em consideração informações reportadas pelos consumidores.

Palavras-chave: Mídias Sociais; Mineração de Texto; Gestão de Relacionamento com Clientes; CRM Social; Reclamações.

ABSTRACT

Digital media are increasingly present in the daily life of human beings. This fact contributes to the increasing volume of user-generated content. From a practical point of view, the analysis of these data requires different perspectives and methods to obtain consistent results. These analyses can support managers' decision making by identifying needs and problems, guiding the process of continuous improvement of products and services. Therefore, this work proposes an analysis of complaints posted on an online complaints platform to identify points that guide companies' decision making and, consequently, improve the relationship with customers. The results obtained allow the identification of problems and their relationships. This study's main contribution is the provision of an approach that helps in the corporation's strategic planning, taking into account situations reported by customers.

Keywords: Topic Modeling, Complaints, eWOM, Text Mining, Social Media.

4.1 INTRODUÇÃO

As mídias sociais oferecem uma ampla gama de funcionalidades que permitem ao usuário *criar* e *compartilhar* conteúdos e o consumo de conteúdo online (CAIR, HAYES, 2015; SIKHA et al., 2018). Este fenômeno impacta diretamente no relacionamento entre empresas e consumidores, isso ocorre devido à popularização da internet e a existência das redes e suas plataformas sociais que, consequentemente, tornam os consumidores mais engajados com internet e na troca de informações sobre produtos e serviços (LOHIO et al., 2017). Consequentemente, percebe-se um aumento substancial na quantidade de conteúdos gerados pelos usuários (User Generated Content - UGC) (BAHAR, MUHAMMAD, 2016; LOHIO et al., 2017; NUSAIR et al., 2017).

UGC pode ser definido como qualquer forma de conteúdo criado, divulgado e consumido por usuários (KIM, JOHNSON, 2016), podendo incluir também dados relacionados a marcas, produtos ou serviços publicados em mídias sociais, os quais constituem uma subcategoria chamada de boca a boca virtual (Electronic Word of Mouth - eWOM) (ALMEIDA, CIRQUEIRA, LOHIO, 2017; SCHMÄH, WILKE, RUSSMANN, 2017). Este cenário impõe um um grande desafio para negócios, pois os usuários não apenas criam e compartilham conteúdo pessoal em seus perfis, mas também, recomendações, opiniões, reclamações e impressões sobre produtos e serviços (ALI, REINHOLD, 2012). eWOM é visto como uma forte determinante na decisão de compra, haja vista que cerca de dois terços dos consumidores verificam as avaliações de produtos, serviços e marcas antes de decidirem adquiri-los (AHMAD, LAROUCHE, 2017; CONSTANTINES, HOLLEKHOVSKY, 2016). A análise de dados relacionados ao eWOM tem o potencial de auxiliar na tomada de decisões por gestores, gerando respostas e melhorias significativas a partir da identificação de necessidades e problemas e serviços envolvidos (GAVILANES, FLATIN, HILDEB, 2018; JONWILLER, STELEN, 2015).

Pesquisas na literatura revelaram que há poucas iniciativas que avaliam a mídia e extração de conhecimento de eWOM expressas em plataformas de reclamações online, focando-se o uso massificado de diversas plataformas de eWOM como fonte de dados para a geração de conhecimento, tal como Facebook (BAHAIL, MITA, 2016; KIM, JOHNSON, 2016; LIU et al., 2017; VERMEER et al., 2019), Twitter (CHAKRABORTY et al., 2017; JONWILLER, STELEN, 2015; VERMEER et al., 2019), e reviews em loja de aplicativos (ALI, JOHANNITE, MORIAN, 2017; MCHURRY et al., 2018; VU et al., 2018).

No Brasil, plataformas de reclamações online são bastante relevantes e influentes. Segundo o site (ALEXA, 2019), no período de 01/11/2019 a 01/02/2020 o Reclame Aqui esteve entre as 25 sites mais acessadas do país, no qual, diariamente cada usuário fez em média três visitas na página com um tempo médio de 10s permanência nas interações. Além disso, percebeu-se que as empresas de telecomunicações são as mais mal avaliadas de acordo com o ranking que o sitebra as 120.000 cadastradas no plataforma. Diante do

contexto apresentado, com destaque para a lacuna na literatura e do potencial de análise e de geração de conhecimento que eWOM apresenta, as seguintes perguntas de pesquisa orientaram o presente trabalho:

- PP1: Quais são os principais tópicos presentes em reclamações envolvendo empresas de telecomunicação?
- PP2: Como esses tópicos estão relacionados entre si?
- PP3: Quais os padrões distributivos das reclamações considerando dimensões geográficas?
- PP4: Quais as implicações práticas do resultado das análises conduzidas para os negócios?

Para responder às perguntas da pesquisa, usou-se dados de reclamações extraídos do site ReclameAqui de quatro empresas do setor de telecomunicações atuantes no Brasil. Este setor foi escolhido devido sua importância na garantia do desenvolvimento de uma sociedade, pois traz consigo a inclusão digital, qualidade de oportunidade, facilidade de transações e comunicação entre indivíduos (FRANKFELT, 1993; THYSSON, 2014; MUTAIRI *et al.*, 2016; SHARMA *et al.*, 2014). Devido às características destes trabalhos, as empresas foram selecionadas considerando o número de clientes e a presença em todo o território nacional, e de acordo com a Agência Nacional de Telecomunicações (ANATEL) (2020), as *Market share* das empresas escolhidas sendo trabalho representam juntas 97% de telefonia móvel, 72% da banda larga fixa, 90,9% de TV por assinatura e 94,0% de telefonia fixa.

O restante do artigo encontra-se organizado como segue. Na Seção 2 são apresentados os trabalhos relacionados. Na Seção 3 a metodologia utilizada é descrita. Os resultados são discutidos na Seção 4. Por fim, as conclusões do estudo e sugestões de trabalhos futuros são apresentadas na Seção 5.

4.2 TRABALHOS RELACIONADOS

Para que o conteúdo de mídia social seja útil para geração de conhecimentos que auxiliem a tomada de decisões, é necessário a definição e uso de metodologias para a extração dos dados. Considerando que alguns dados são de difícil acesso, Sfriso *et al.* (2010) define uma arquitetura para um framework com desafios e vantagens a serem seguidas para a extração de dados. Com isso, apresenta-se como um *Webcrawler* pode ser empregado para isso, no processo de extrair e identificar grandes quantidades de conteúdo gerado pelo usuário. Além disso, apresenta a importância das ciências sociais e

da computação no processo de análise de dados sociais. Estudos semelhantes baseiam-se neste framework, a citar (NETTO et al., 2019; RODRIGUES JR; LOBATO, 2019).

Devido à indistintabilidade das léxias e a necessidade de ajustes gramaticais, o pré-processamento se faz um processo fundamental para a análise de conteúdo de mídias sociais para se garantir a confiabilidade das conclusões. Chiquerra et al. (2018) realizaram uma análise na literatura para identificar os principais métodos utilizados para tratar conteúdos escritos em Português-Brasil. Foram reunidos um total de 12 artigos relevantes, os quais possibilitaram a listagem dos principais métodos e etapas necessárias.

Espetando aos tipos de análises, a metodologia de tópicos como destaque, uma vez que permite a identificação de certos padrões nos dados, nos quais seriam difícil a descoberta manual. Para isso, Tunda et al. (2018) utilizaram técnicas de metodologia de tópicos para identificar o nível de engajamento de usuários. Foram reunidos mais de 1,9 milhões de tweets de 146 usuários e, a partir das análises desses dados, os autores demonstraram que menções, apelo emocional e discussão em torno da vida pessoal são fatores prediletos de um usuário em que é possível a divulgação dos mais variadas tipos. Outros trabalhos apresentam a metodologia de tópicos como uma forma de extrair os termos importantes no texto, tal como (CHIQUEIRA et al., 2017; ALJHOU AN; JANSEN, 2019; ALMEIDA; CHIQUEIRA; LOBATO, 2017; Li et al., 2019).

Sob o ponto de vista da análise de eWOM, Rowe e Yang (2015) apresenta uma análise do setor de turismo a partir de reviews publicadas em plataformas online. Os autores analisaram seis aspectos relevantes em hotéis, com cinco tipos de viagens pré-definidos e visitantes de dois países diferentes. Como resultado, os autores verificaram que o “valor” e o tipo do “quarto” são os elementos que mais influenciaram em uma avaliação. Assim sendo, a análise de conteúdos de mídias sociais permite que as ações nestas plataformas sejam mais eficazes e eficientes (CHITT MALLA; OGHARI; PARIDA, 2018).

No trabalho de Tang e Guo (2015) destacam a viabilidade e eficácia do uso de técnicas de mineração de textos para prevenir o conteúdo. No mesmo sentido, Calmeida et al. (2019) apresenta a técnica de mineração de textos como forma eficaz para a descoberta de embarralhamento e conteúdos sensíveis. Os autores demonstram a identificação das palavras mais relevantes em conjuntos textuais e esse potencial de realizar previsões futuras com base nos resultados.

Para o desenvolvimento deste trabalho optou-se pela metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM). A metodologia CRISP-DM é implementada a partir de um processo hierárquico, consistindo em um conjunto de tarefas que descrevem quatro níveis de abstração (CHINCHOLA; PEREIRA, 2016). Considerando os níveis de abstração propostos, os trabalhos de Hollins (2015), Scholer et al. (2019), Ward (2000) apresentam descrições detalhadas dos processos que compõe o CRISP-DM. Estes processos formam o ciclo de um projeto de mineração de dados composto de seis

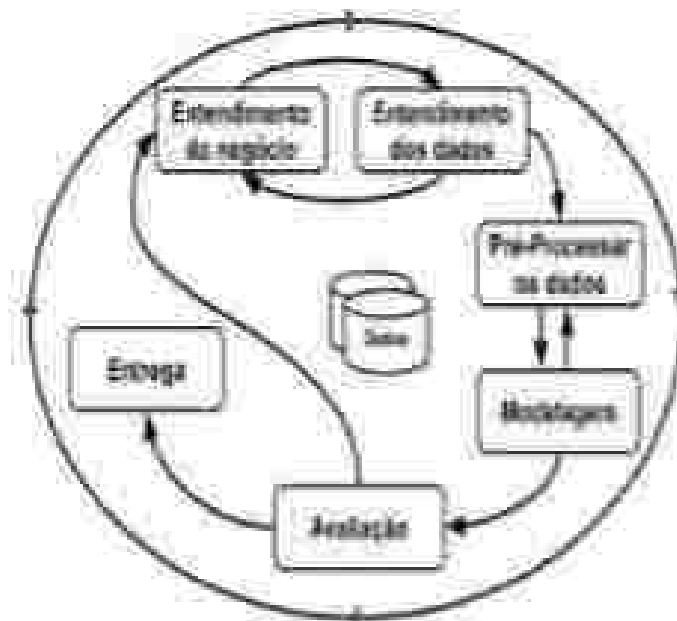


Figura 7: Diagrama do desenvolvimento do modelo CRISP-DM (Adaptado de [Wirth \(2000\)](#)).

base, conforme pode ser observado na Figura 7.

4.3 METODOLOGIA

Para o desenvolvimento deste trabalho optou-se pela metodologia *ross Industry Standard Process for Data Mining* (CRISP-DM). A metodologia CRISP-DM é implementada a partir de um processo hierárquico, estruturado em um conjunto de tarefas que descrevem quatro níveis de abstração ([CHENCHILLA, FERREIRA, 2010](#)). Considerando os níveis de abstração propostos, os resultados de [Rialho \(2015\)](#), [Schäfer et al. \(2015\)](#), [Wirth \(2000\)](#) apresentam descrições detalhadas dos processos que compõem o CRISP-DM. Estes processos formam o ciclo de um projeto de mineração de dados composto de seis bases, conforme pode ser observado na Figura 7.

Devido a características específicas deste método, o processo de mineração de dados não é finalizado quando uma solução é implementada. Nesse caso, as lições aprendidas durante cada etapa podem gerar novas possibilidades de análises e novos resultados ([WIRTH, 2000](#)). Nas próximas subseções são descritas as etapas de Entendimento do Negócio, Entendimento dos Dados, Pré-Processamento dos Dados e Modelagem. A etapa de Avaliação é descrita na Seção 5 e a Entrega é feita por meio da apresentação dos resultados aos stakeholders.

4.3.1 ENTENDIMENTO DO NEGÓCIO

Partindo em consideração a realidade de que a metodologia CRISP-DM possibilita, nessa etapa o entendimento o entendimento do contexto em que se analisam os dados.

ser aplicadas, a partir da definição dos objetivos da mineração de dados. Neste contexto, verifica-se que as reclamações podem representar uma importante fonte para obtenção de uma grande quantidade de informações valiosas sobre produtos e serviços feitos de forma voluntária. Com base nisso, os seguintes objetivos foram definidos:

- Identificar os principais assuntos/problemas reportados nas reclamações;
- Identificar aspectos específicos das reclamações;
- Analisar a distribuição das reclamações considerando dimensões geo-linguísticas;
- Identificar as implicações práticas das reclamações nas empresas.

4.3.2 ENTENDIMENTO DOS DADOS

Nesta etapa são realizadas a coleta, descrição, exploração e verificação da qualidade dos dados extraídos. Devido as dificuldades de mineração da API brasileira, os dados foram extraídos da plataforma de reclamações online chamada Reclame Aqui³ por meio de um WebCrawler escrito em Python, o qual utiliza a biblioteca *requests*⁴. Esta plataforma foi escolhida devido sua popularidade. Dentre todos os sites acessados no Brasil, o ReclameAqui é o 2º site com maior número de acessos (ALEXA, 2019), sendo o site mais popular no segmento de reclamações.

Como alvo da extração, foram selecionadas as quatro maiores empresas do setor de telecomunicações brasileiras. Estas empresas foram escolhidas devido a abrangência nacional de prestação de serviço, as quais atendem milhões de clientes em todas as regiões do país. Além disso, essas empresas são apontadas como as piores empresas no ranking brasileiro pelo ReclameAqui, o qual contém mais de 120.000 empresas cadastradas. Cada reclamação extraída era composta pelas dados detalhados na Tabela 7.

Campo	Descrição
Empresa (E)	Identificação da empresa relacionada na reclamação
Reclamação (R)	Identificação da reclamação
Título da Reclamação	Título dado pelo usuário a reclamação
Reclamação	Relato do problema reportado
Usuário e cidade	Usuário e cidade do consumidor que relatou a reclamação
Data/hora	Data e horário do relato

Tabela 7: Descrição dos dados coletados

A fim de verificar a qualidade dos dados extraídos, foi realizada uma comparação manual com os dados da plataforma. Para isso, foi utilizada uma constante binomial que

³ *Disclaimer: É importante ressaltar que todos os dados coletados no ReclameAqui serão utilizados exclusivamente para fins de prova de conceito. Os dados não tem qualquer impacto no seu cotidiano desses dados.*

⁴ <https://requests.readthedocs.io/en/master/>

representa um grau de confiança de 95% e uma margem de erro de 4% considerando o total de reclamações coletadas na plataforma por este estudo. Por fim, foi verificado que, com o uso da ferramenta de extração, os dados mantiveram o padrão de qualidade observado no site da plataforma.

4.1.3 PRÉ-PROCESSAMENTO DOS DADOS

Nesta etapa foi realizada a aplicação de técnicas de pré-processamento nos textos de cada reclamação. Em todos os dados de reclamações foram realizadas as seguintes tarefas de remoção: de pontuação, de URLs, abreviatura, acentuação e de caracteres especiais (CHIQUEIRA et al., 2018).

A remoção de pontuação e de URLs significa que todos conjuntos de caracteres que representa uma pontuação (e.g. "OK", "O") ou um endereço de algum site (e.g. "www.site.com.br") foi removido. Da mesma forma, acentuação, acentuação e caracteres especiais foram retirados, visto que não eram necessários para as análises. Palavras que não consideravam stopwords, ou seja, palavras que não contribuem para o significado do texto (e.g. "e", "do", "em"), foram eliminadas.

4.1.4 MODELAGEM

Esta etapa foi dividida em duas fases: a aplicação da extração de tópicos e a correlação entre os tópicos encontrados. Na primeira fase de modelagem foram aplicadas técnicas de extração de termos relevantes sobre os dados. Alguns algoritmos, tal como o *Non-Negative Matrix Factorization* (NMF), *Latent Dirichlet Allocation* (LDA) e *Latent Semantic Analysis* (LSA) foram utilizados nesta fase.

Durante o processo de análise qualitativa e categorização dos tópicos percebeu-se que os termos obtidos pelo NMF estavam muito relacionados entre si e que representavam tópicos muito coerentes e diversos (CHEN et al., 2018). Devido a isso, o *Term Frequency-Inverse Document Frequency* (TF-IDF) juntamente com o NMF foram adotados neste trabalho com o objetivo de classificar todos os palavras por ordem de importância no conjunto de textos, tal como demonstrado por Salimian et al. (2018), Todorjak, Mikic e Dondic (2014).

A modelagem e análise de todos as reclamações foi realizada de acordo com a empresa sob uma perspectiva nacional e regional. Os tópicos foram desenvolvidos pelos autores de forma manual com base nos termos obtidos na modelagem. Na segunda fase foi realizado o estudo de correlação dos tópicos obtidos. Vale destacar que um tópico representa um conjunto de termos relacionados, sendo que cada termo é associado há um conjunto de reclamações. Neste caso, para realizar a correlação, os dados foram modelados da seguinte forma: cada tópico é representado por um nó, as reclamações foram convertidas em arestas, as quais ligam os diferentes tópicos os quais tem correlação.

Neste sentido, o conjunto de dados foi então transformado em registros contendo pares de arestas, por meio de uma combinação simples denotada por:

$$r = \frac{n!}{2(n-p)!} \quad (4.1)$$

Sendo que r é a quantidade de registros resultantes, n é o número de palavras-chave do traduíng, e p foi definido como 2 (dois), pois as arestas são formadas aos pares. Como resultado desta modelagem, foi criado um arquivo *Comma-Separated-Values (CSV)*, a qual que pode ser visualizada com o suporte de uma ferramenta de análise de redes. Na presente estudo o software *Cyphi*³ foi utilizado para este fim.

4.4 RESULTADOS

Nesta seção serão apresentados os resultados obtidos a partir das análises descritas anteriormente. Os resultados são divididos em três subseções. A primeira apresenta os dados coletados e a relação do número das reclamações com a distribuição da população brasileira. A modelagem de redes em termos nacional e regional é realizada na segunda subseção. E, por fim, na terceira subseção são apresentadas as análises geo-temporal das reclamações obtidas.

4.4.1 EXTRAÇÃO DOS DADOS

O processo de coleta dos dados resultou em um total de 397.950 reclamações. Considerando a abrangência do setor de telecomunicação e sua importância nacional (STONIC, 2018), foram utilizadas apenas reclamações do período de 01/01/2018 a 01/09/2019, resultando em 225.503 reclamações. Na Tabela 4 é apresentada as quantidades de reclamações coletadas em nível regional e nacional de cada empresa.

Tabela 4 – Distribuição dos dados pelas regiões do país

	Norte	Nordeste	Centro-Oeste	Sul	Sudeste	Total
Tim	1.377	10.090	1.726	47.229	9.292	71.714
Vivo	1.114	8.956	4.810	68.829	9.294	93.003
Claro	1.257	6.863	2.763	19.458	4.779	35.110
Oi	974	4.801	2.752	27.058	1.732	37.317

O Brasil possui aproximadamente 210 milhões de habitantes distribuídos em 5 regiões (IBGE, 2019). As reclamações analisadas foram redistribuídas de acordo com a organização e a respectiva região de origem, podendo ser observado na Tabela 4. A região Sudeste tem um destaque frente as demais. Esta região possui aproximadamente

³ <https://cyphi.org/>

Tabela 9 Número de tópicos encontrados e resultados

	Tópicos encontrados	Tópicos resultante
Tm	31	17
Vivo	29	20
Oi	26	17
Claro	26	13

42% da população brasileira e apresenta uma taxa superior a 60% do total de reclamações criadas. Este fato pode ser justificado, uma vez que a região sulista é a maior fica do país, com aproximadamente 53% de todo o Produto Interno Bruto (PIB) produzido. Enquanto, por exemplo, a região Norte representa aproximadamente 6% do total (IBGE - Instituto Brasileiro de Geografia e Estatística, 2020).

4.4.2 - MÓDULO DE TÓPICOS

Nesta subseção é apresentada a modelagem de tópicos sob duas perspectivas: 1) tópicos de reclamações de todo o país; 2) a partir de uma distribuição regional. Devido às características da plataforma analisada, na qual o usuário seleciona a categoria do problema, optou-se por utilizar a mesma quantidade de categorias de problemas como a quantidade de tópicos a serem modelados. Isto permite dimensionar o espaço de busca. Maiores detalhes são discutidos a seguir.

4.4.2.1 - PANORAMA NACIONAL

O total de reclamações de cada empresa (apresentadas na Tabela 8) foram utilizadas para montar os tópicos de reclamações no panorama nacional. A partir da limitação do número de tópicos frente as categorias de problemas da plataforma, obtiveram-se a seguinte distribuição por empresa: 31 tópicos da Tm; 29 na Vivo; 26 tópicos na Oi e 26 na Claro. Após uma análise inicial, alguns tópicos semelhantes (termos similares, por exemplo) foram eliminados. Além disso, foram consideradas apenas tópicos únicos e robustos. O resultado encontra-se sintetizado na Tabela 9.

Na Tabela 10 são expostos os principais tópicos das reclamações de acordo com a quantidade observados na Tabela 3. Ao observar os dados é possível reconhecer os problemas específicos de cada empresa no país. Tais dados podem ser utilizados como fundamento para o início de um processo de melhoria do serviço, fidelizar clientes e conquistar vantagens competitivas em relação aos demais concorrentes (CAVALANES, FLAUTIN, MORTOLA, 2018; FINWILLER, STOLIS, 2017).

Para aprimorar a descrição dos resultados, alguns detalhes sobre o funcionamento da análise foram se necessários, isto é, características inerentes da empresa analisada. Por exemplo, a Tm oferece um plano exclusivo para seus clientes, o qual é possível acessar

por meio de um canal de contato por outros canais desse plano⁴. A partir da identificação do tópico relacionado é possível verificar que reclamações relacionadas a este serviço são frequentes e podem ser relacionadas este processo de negócio utilizado pela empresa.

Tabela 10 Modelagem de tópicos para análise do problema relatado

Top	Vivo	Oi	Claro
Atendimento	Atendimento	Atendimento	Atendimento
Atracção chip	Atracção chip	Atracção linha	Atracção linha
Cancelamento	Cancelamento	Cancelar serviço	Cancelamento
Cobrança incorreta	Cobrança incorreta	Cancelamento	Cobrança incorreta
Cobrança indevida	Cobrança indevida	Cobrança indevida	Cobrança indevida
Cobrança indevida	Internet malfuncionando	Indicação de equipamento	Internet malfuncionando
Conversão para plano	Ligação telemarketing	Internet malfuncionando	Ligação telemarketing
Franquia de dados	Linha cancelada	Ligação telemarketing	Linha cancelada
Ligação telemarketing	Linha fixa	Linha cancelada	Linha fixa
Linha fixa	Mudança de endereço	Linha fixa	Multa por fidelidade
Nome registrado	Mudança de plano	Multa por fidelidade	Nome registrado
Linha cancelada	Multa por fidelidade	Nome registrado	Nome cancelado
Pagamento	Nome registrado	Plano controle	Pacotes de dados
Plano	Nome cancelado	Plano	Pagamento
Plano Controle	Pagamento	Portabilidade	Plano
Portabilidade	Portabilidade	Qualidade de serviço	Portabilidade
Recarga	Qualidade de serviço	Recarga	Recarga
Valor serviço	Recarga	Teléfono fixo	Sinal de aviso
	Visita técnica		Serviço de técnico

De forma geral, as reclamações concentram-se relacionadas a falta de um tópico. A correlação entre os tópicos das reclamações das comunidades aponta a verificação da distribuição desta rede de instalações. Na Figura 8, são expostas as relações entre os tópicos, sendo que tanto o tamanho das palavras quanto a repetição e variação das ligações representam a importância e o peso da relação.

Após analisar os dados presentes Figura 8, são pontos notáveis destaques em relação às comunidades estudadas:

- **Top** - Justa como seu principal problema o atendimento. Na Figura 8a é visível que há associações entre diversos tópicos o que indica que a empresa tem dificuldade de oferecer soluções para seus clientes por meio do atendimento. Além disso, problemas como 'pagamento', 'cobrança indevida' e 'recarga' que estão fortemente associados e indica falhas recorrentes na maneira que a empresa trata o processo de cobrança dos clientes.
- **Vivo** e **Oi** argumentam como seus principais gargalos o atendimento e a cobrança indevida. Na Figura 8b (Vivo) e na Figura 8c (Oi) há uma grande associação entre os dois problemas, sendo que para a empresa Vivo o atendimento tem peso maior.

⁴ <https://www.timbata.com.br/timbata/contato-ou-ajuda>

Para a empresa Oi, o tópico “conexão” é um problema específico das regiões Sudeste, Nordeste e Centro Oeste. Por fim, a Claro não apresenta problemas específicos regionais. Contudo, alguns problemas nacionais que não fazem parte dos contextos regionais, como é o caso da região Sudeste que não apresenta dentro os principais tipos de problemas o “número excedido” e o “perda de dados”.

Diante disso, empresas de telefonia podem analisar quais os tópicos mais relevantes em uma determinada região, bem como os tópicos que não são relevantes (ou período estudado) para os clientes em cada área do país. Tal informação permite que áreas menos populosas e com poucas reclamações sejam tratadas de acordo com a sua especificidade, eliminando a possível visão que as áreas mais populosas podem causar em análises que consideram reclamações de todo país e elevam-se somente problemas nacionais.

É importante ressaltar que, a análise de relação entre o tipo de reclamação categorizada na subseção 4.4.2.1, é perfeitamente aplicável aos contextos regionais, estaduais e municipais, no entanto por questões de espaço estas análises foram expostas no presente trabalho.

Tabela 11: Tópicos por região das empresas estudadas.

	Rio de Janeiro	Sul	Pernambuco	Paraná	Centro Oeste
Oi	• acesso à internet • custo por minuto • qualidade de serviço • custo de ligação	• internet celular • acesso à internet • custo de ligação	• acesso à internet • custo por minuto • acesso de internet	• qualidade de ligação • custo por minuto • qualidade • acesso de internet • custo de ligação	• internet celular • custo por minuto • qualidade de serviço • acesso de internet
	• plano celular	• internet celular • plano celular	• comunicação • plano celular	• voz, SMS • internet, qualidade • plano celular	• comunicação • plano celular
Vivo	• custo de ligação • qualidade de comunicação • acesso • custo de ligação		• plano	• acesso de internet	• comunicação
	• internet móvel • acesso, internet • plano, celular • qualidade de comunicação • custo por minuto • qualidade de internet • qualidade de serviço • internet, fixo • custo de ligação	• internet, plano • internet	• comunicação • qualidade de internet • internet, fixo	• qualidade de internet • qualidade de internet • internet • plano	• internet, plano celular • internet, comunicação • internet, fixo
Claro	• internet • comunicação	• comunicação • comunicação	• comunicação	• internet • comunicação • comunicação	• comunicação
	• custo por minuto • plano celular • internet, fixo	• comunicação • plano celular	• plano celular • internet, fixo	• internet, acesso • internet, comunicação • plano celular • qualidade de serviço • internet, fixo	• comunicação • plano celular • qualidade de serviço • internet, fixo
Claro	• comunicação • plano celular • voz, SMS • acesso de internet • acesso de internet	• plano celular • plano celular • plano celular • acesso de internet • acesso de internet	• plano celular • plano celular • plano celular • plano celular • plano celular	• comunicação • plano celular • plano celular • plano celular • plano celular	• plano celular • plano celular • plano celular • plano celular • plano celular

4.4.2.3 DISTRIBUIÇÃO “CDO-TEMPORAL”

As reclamações coletadas serão distribuídas em todo território do Brasil. São distribuídas de forma heterogênea e desproporcional a taxa populacional de cada região. Na Tabela 12 é possível observar como se dá a distribuição geográfica das reclamações ao país através da computação das taxas populacionais, PIB, Linhas aéreas e de reclamações em cada região. Os dados são apresentados de forma a relacionar ao PIB, taxa populacional e de linhas aéreas aéreas em cada região.

Tabela 12 – Proporção Populacional (IBGE, 2019a), PIB (IBGE – Instituto Brasileiro de Geografia e Estatística, 2020), Linhas Aéreas (ANAC, 2020) e de reclamações por Região

	População	PIB	Linhas aéreas	Reclamações
Norte	9%	6%	7%	2%
Nordeste	27%	10%	24%	12%
Centro-Oeste	14%	17%	8%	6%
Sudeste	42%	52%	46%	49%
Sul	14%	14%	15%	12%

Na Figura 9, são expostas as quantidades de reclamações por horário e dia da semana, sendo que na Figura 9a contém a distribuição de horas as reclamações em 24 horas e na Figura 9b contém a distribuição de horas as reclamações nos dias da semana.

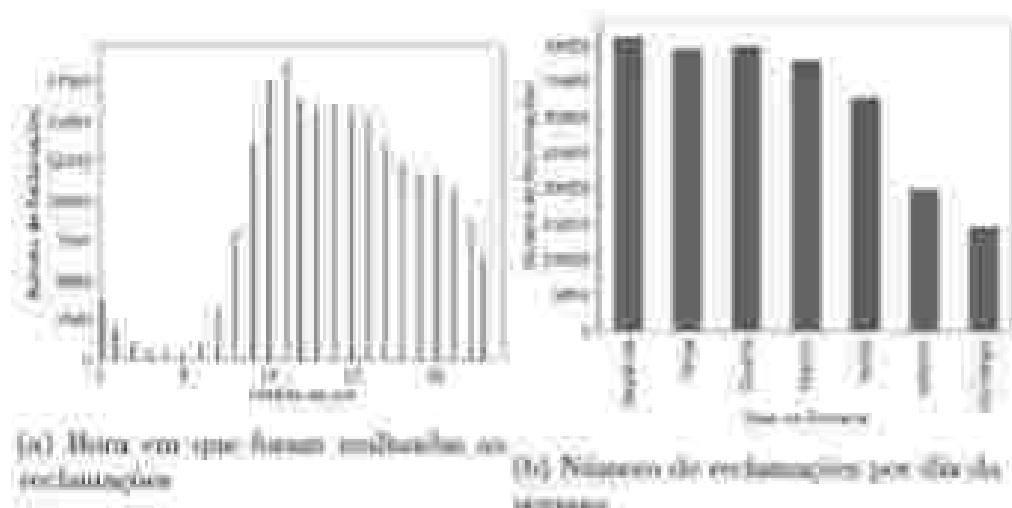


Figura 9 – Número de reclamações por dia da semana.

Observando-se a Figura 9 é possível perceber que a distribuição em relação aos dias da semana e horários seguem também padrões com variações para horas não finais de semana e entre outras horas da noite e até horas do dia seguinte. No entanto, essas informações podem auxiliar as empresas na definição de horários e presença humana para solucionar os problemas levantados pelos plataformas online de reclamações de forma mais eficiente.

A análise dos dados disponíveis na Tabela 12 e figura 8 nos revela alguns insights importantes sobre a distribuição das reclamações no país, a saber:

- Região Norte é a região com menor PIB do país, menor número de habitantes e em quantidade de linhas de celulares ativos. Há uma certa discrepância em relação à proporção de reclamações de outras regiões. Talvez as cidades tenham a taxa de reclamação mais próxima do número de linhas ativas.
- Região Centro-Oeste tem cerca de 8% do total de linhas ativas e 6% das reclamações coletadas são referentes a essa região.
- Região Sudeste tem a maior taxa populacional e o maior número de reclamações no país. Sendo assim, a taxa de reclamação desta região é bastante elevada, ficando cerca de 21 pontos percentuais acima do número de linhas ativas.
- Todas as outras regiões, exceto a região sulista, possuem taxas de reclamações inferiores a taxa de linhas ativas. O que pode indicar que se chama de região sulista são mais propensas a utilizar plataformas de reclamações online para avaliar suas queixas.

4.5 CONSIDERAÇÕES FINAIS

Neste artigo, foram analisadas reclamações de quatro empresas do setor de telecomunicações brasileiro identificando estruturas quais eram os principais tópicos nas reclamações, como elas se relacionam entre si e qual a sua distribuição geo-temporal. Para tal, utilizou-se a tecnologia CRISP-DM, sendo que as fase de extração de conhecimento foram aplicadas métodos de modelagem de tópicos a nível nacional e regional. Os resultados mostram que o uso de técnicas de extração de texto combinado com uma análise geo-temporal permite a identificação e a correção de problemas, a análise de grandes volumes textuais de forma rápida e análises estratégicas, além de prover bases sólidas para a tomada de decisão por gestores.

As análises conduzidas permitem responder as perguntas de pesquisa, relevando importantes insights sobre as empresas analisadas. Por exemplo, em relação à P11 verificou-se que dentre os principais tópicos das reclamações o “atendimento” e “cobrança indevida” são os tópicos que mais se destacam nas empresas analisadas. Para P12, identificou-se que a relação entre os tópicos é diferente em cada empresa e indicando que a análise de problemas e ações solucionadoras variam entre os consumidores. Acerca da P13, verificou-se que existe uma diferença regional em relação aos tópicos das reclamações analisadas, porém não foram observadas diferenças significativas no número de reclamações durante o ano. Por fim, por meio da resposta da P14, os resultados auxiliam as empresas de telecomunicações no processo de tomada de decisão, uma vez que essas empresas podem

direcionar seus esforços para solucionar os problemas de seus clientes de acordo com o contexto em cada região.

Diante disto, a principal contribuição deste estudo está na produção de uma abordagem de análise de reclamações que identifique as suas necessidades das análises de telefonia, auxiliando as empresas identificadas no estudo na implementação de soluções personalizadas por tópico e por região.

No entanto, este estudo apresenta algumas limitações que precisam ser trazidas em trabalhos futuros. A primeira está relacionada a forma como os tópicos foram determinados os quais podem conter viés dos autores. A segunda limitação é relacionada às análises e resultados obtidos, não não incluindo a aplicação prática dos resultados nas empresas. Como trabalhos futuros pretendemos incluir a validação cruzada em todas as etapas da metodologia, e expandir as análises incluindo o uso prático em empresa dos resultados encontrados.

5 ARTIGO - ANÁLISE COMPARATIVA DAS PRINCIPAIS PLATAFORMAS DE RECLAMAÇÕES ONLINE: IMPLICAÇÕES PARA ANÁLISE DE MÍDIA SOCIAL EM NEGÓCIOS

ABSTRACT

New forms of relationships between companies and customers have been introduced through the massive use of social media, and have transformed the way in which customers make purchasing decisions. This new reality emphasized the importance of content analysis related to brands and products published by consumers on social media platforms. In this sense, this article presents a comparative analysis of the two largest online complaint platforms in Brazil, Reclame Aqui and Consumidor.gov. The analysis was conducted from textual data mining of consumer complaints from five major companies in the Brazilian e-commerce sector, in order to provide a basis for understanding the challenges and opportunities in the analysis of these social media in business. The results show that the way in which companies operate on these platforms must be specific to each platform, since each platform has its particularities, such as consumer groups, content and main types of problems.

Keywords: Social Media, Data Mining, Topic Modeling.

RESUMO

Novas formas de estabelecimento entre empresas e clientes foram introduzidas através de uma mudança de mídia social, e isso transformou a forma como a qual clientes tomam decisões de compra. Esta seção estabelece explicitamente a importância da análise do conteúdo relacionando a natureza e produtos publicados por consumidores em plataformas de mídia social. Neste sentido, este artigo apresenta uma análise comparativa das duas maiores plataformas de recomendações online no Brasil, o *Reclame Aqui* e o *Consumidor.gov*. Nossa análise não utiliza dados em conteúdos provenientes de menções de dados textuais de recomendações de cinco grandes empresas do setor de comércio eletrônico, como o objetivo de fornecer uma base para compreender os desafios e oportunidades aos analistas de mídia social em negócios. Os resultados mostram que a forma de atuação das empresas através plataformas deve ser equilibrada para cada plataforma, pois existem particularidades, tais como grupo de comunidades, conteúdos e principais tipos de problemas.

Palavras-chave: Mídia Social, Menções de Dados, Metodologia de TDA

5.1 INTRODUÇÃO

O surgimento de novas ferramentas de comunicação modificaram a forma com a qual as pessoas tomam suas decisões de compra (OBENCA-ROCHA; CHALMEIDA, 2016; LEBLANC *et al.*, 2017). O cenário de compartilhamento de informações online facilitou o diálogo entre consumidores de todo o mundo, independentemente de suas preferências geográficas (CONSTANTINIDES; HOLLETSCHOVSKY, 2016). E, assim como o boca-a-boca tradicional se faz presente na sociedade, o ambiente virtual tornou-se um dos meios de disseminação e troca de experiências a respeito de produtos e serviços (KIM; JOHNSON, 2010). Como consequência, surgiu o conceito de boca-a-boca virtual (*Electronic Word-of-Mouth*, eWoM), definido como o ato de compartilhar opiniões sobre produtos e serviços em mídias online (SCHMIDT; WILKE; HEESMANN, 2017).

O eWoM tem se mostrado ainda mais eficaz que o "boca-a-boca" tradicional, devido a sua facilidade de propagação e de engajamento (CONSTANTINIDES; HOLLETSCHOVSKY, 2016). Por este motivo, as organizações passaram a ser chamadas de *prosumers*, uma vez que são responsáveis tanto pelo consumo quanto pela produção de conteúdo (ROY; DATTA; MUKHERJEE, 2019). Ao estudar conteúdos de eWoM, as empresas têm uma rica fonte de conhecimento sobre seus clientes e quais são suas preferências (ALDOUS; AN; JANSEN, 2019). Porém, a obtenção de conteúdo tanto o entendimento individual dos clientes um grande desafio (ROBERTO *et al.*, 2017).

Na literatura, há diversos estudos que utilizam dados de reviews e de relações online obtidos de diferentes plataformas de mídias sociais tais como Twitter, Facebook, TripAdvisor, Booking.com e entre outras (DAFTAR MUDA, 2016; SILVA *et al.*, 2017; XIANG *et al.*, 2017; VERMEER *et al.*, 2019). Geralmente, estes estudos utilizam em suas análises um conjunto amostral de dados para extrair informações que permitissem a detecção, descrição ou previsão de padrões que influenciam na tomada de decisões *teóricas e/ou práticas*. Entretanto, há poucos estudos tal como o de Almeida, Ladeira e Junior (2015) que utilizam em suas análises plataformas específicas para publicações de reclamações online de *padrões e/ou práticas*.

Com isto em mente, este estudo analisa comparativamente o conteúdo textual das reclamações de duas plataformas de reclamações online¹, a saber a Reclame Aqui (RA) e a Comunidade.gov (CGOV). O objetivo é prover uma base para a compreensão das desafios e oportunidades que estas plataformas proporcionam para o desenvolvimento de análises de mídias sociais que visem a tomada de decisões em diversas segmentos da sociedade.

O restante do artigo encontra-se organizado como segue. Na Seção 5.2 são apresentadas as tradições relacionadas, na Seção 5.3 é descritos processos de condução da pesquisa,

¹ *Observação:* É importante ressaltar que todos os dados utilizados no Reclame Aqui e na Comunidade.gov serão utilizados exclusivamente para fins de prova de conceito. Os autores não têm qualquer intenção de uso comercial desses dados.

na Seção 5.4 são apresentadas as resultados decorrentes das análises e por fim na Seção 5.5 o autor apresenta suas considerações finais no desenvolvimento do estudo.

5.2 TRABALHOS RELACIONADOS

O engajamento de comunidades em múltiplas plataformas de mídia social é um grande desafio que empresas têm enfrentado. Neste sentido, [Adams, An e Jansen \(2019\)](#) buscou entender como o engajamento de usuários ocorre em diferentes plataformas de mídia social e como isto afeta o conteúdo produzido, através da modelagem de tópicos em dados de publicações, comentários e curtidas extraídos de perfis de empresas no *Facebook*, *Instagram*, *Twitter*, *YouTube* e *Reddit*.

[Quattione et al. \(2018\)](#) propõe a junção de vários métodos para analisar as reviews em plataformas de economia compartilhada, com o objetivo de gerar conhecimento e valor para o setor por meio da classificação e análise semântica de milhares de reviews. E ao analisar mídia social em busca de padrões que determinam o comportamento do usuário, [Vedharam et al. \(2018\)](#) examinou a viabilidade dos uso dos dados textuais presentes no *Twitter* para determinar padrões alimentícios de usuários. Através da coleta e classificação de dados textuais relacionados a alimentação, verificaram que é possível utilizar as mídias sociais para determinar tendências, tal como determinar o nível de consumo de alimentos saudáveis e não saudáveis. Outro ponto que merece destaque quanto a busca de padrões é o apresentado por [Yang et al. \(2016a\)](#), que estudaram os fatores que influenciam na percepção de valor em reviews. Através de análises de perspectiva dos textos e de seus autores, verificou-se que a legibilidade e sentimento do texto exerce influência significativa na importância que os leitores atribuem a uma review.

Há diferentes tipos de plataformas de mídia social com diferentes formas de interação. [Xiang et al. \(2017\)](#) propõem um modelo que examina comparativamente três plataformas de reviews de consumidores em termos de qualidade das informações relacionadas às avaliações online nesse sites. O objetivo deste trabalho foi fornecer uma base para a compreensão dos desafios metodológicos e para identificar oportunidades para o desenvolvimento da hotelaria e turismo. Foram extraídos 1.431 reviews em três das principais plataformas online no mundo da área, a saber: *TripAdvisor*, *Expedia* e o *Yelp*. Com as análises das características dos dados, foi verificado que, apesar de grande parte da literatura apresentar os textos como fontes primárias de conhecimento para pesquisadores, há, no entanto, diferenças significativas nos reviews em diferentes plataformas.

Assim como [Adams, An e Jansen \(2019\)](#), este trabalho apresenta métodos de mineração de textos providos de mídia social, juntamente com classificações e análises semânticas presentes em [Quattione et al. \(2018\)](#). E em concordância com o trabalho de [Yang et al. \(2016a\)](#), também são verificados fatores que influenciam na percepção de valor

dos reviews, além de medir a viabilidade de determinação de padrões e tendências a partir da coleta e classificação de dados textuais originados de mídias sociais (VYJISWARAN *et al.*, 2015). E os resultados corroboram com a pesquisa de Xiang *et al.* (2017) que observou a partir de características dos dados de reviews que existia diferença notável entre as plataformas de clientes.

5.3 MATERIAIS E MÉTODOS

Nesta seção são apresentados os processos de coleta e análise de dados, baseados nos trabalhos de Xiang *et al.* (2017), Hung *et al.* (2016a). Os trabalhos foram escolhidos devido sua relevância em publicações na área de mineração de dados textuais. Na primeira etapa, os dados de reclamações foram coletados. Posteriormente, foram pré-processados e extraídas as mídias textuais. Por fim, foram construídas as estruturas de coerência e de modelagem de tópicos.

5.3.1 COLETA DE DADOS E PRÉ-PROCESSAMENTO

Devido à sua importância econômica, foi escolhido para a mineração dos dados o setor de comércio brasileiro (ABICOMM, 2019). Neste caso, selecionou-se as cinco empresas com maiores números de vendas em sites lojas virtuais (NETRICA, 2016). Os dados coletados incluem nome da empresa, relato da reclamação, tempo de resposta, estado, cidade e data da reclamação.

Todos os dados de reclamações foram pré-processados utilizando os procedimentos de remoção de pontuação, de URLs, stopwords, números, acentuação e caracteres especiais (CHIRQUEIRA *et al.*, 2019). A remoção de pontuação e de URLs significa que toda string que representa uma pontuação (e.g. "064", "06") ou um endereço de algum site (e.g. "www.abicomm.br") foram removidos. Da mesma forma, ocorreu com números, acentuações e caracteres especiais, visto que são desnecessários para as análises. Palavras que são consideradas stopwords, que não contribuem para o significado do texto, também, foram removidas (e.g. "e", "de", "em"). Esta processo é importante, uma vez que garante uma maior qualidade dos dados para executar as análises propostas.

5.3.2 EXTRAÇÃO DE CARACTERÍSTICAS TEXTUAIS

As características textuais são importantes para detectar padrões que permitam a identificação e comparação de padrões de textos em diferentes locais (HUNZEL *et al.*, 2017). Os dados textuais das reclamações foram processados e analisados, levando em consideração a quantidade de caracteres de cada reclamação e da grau de legibilidade através do Flesch Readability Base Score (FRES). O FRES é uma fórmula desenvolvida por Rudolf Flesch em 1948 utilizada para calcular o grau de legibilidade de um texto (FLESCH, 1960). O

valor de FRES é obtido por meio da seguinte fórmula (OTHMAN et al., 2012):

$$FRES = 396,88 - (1,015 \times ASL) - (81,6 \times ASW) \quad (5.1)$$

A variável *ASL* (*Average Sentence Length*, em português "tamanho médio da sentença") é obtida a partir da divisão do número de palavras pelo número de sentenças. O *ASW* (*Average number of Syllables per Word*) é obtido a partir do número médio de sílabas pelo número de palavras. Os valores de FRES representam o grau de legibilidade de um texto e pode variar dentro dos intervalos de zero a cem, sendo que quanto maior for o valor mais fácil é a leitura de um texto (OTHMAN et al., 2012). Esses valores são classificados de acordo com a Tabela 13, construída a partir dos trabalhos de (PORTO et al., 2011; OTHMAN et al., 2012; HIRSH et al., 2017).

Tabela 13 - Classificação dos graus de legibilidade

Classificação	Grau
1 ^o ao 12 ^o ano	75 a 100
12 ^o ao 14 ^o ano	50 a 75
Ensino Médio e Nível Superior	25 a 50
Textos Acadêmicos	0 a 25

5.1.3 COERÊNCIA E MODELAGEM DE TÓPICOS

A tarefa de determinar o melhor número de tópicos em modelagens de tópicos tem se mostrado muito subjetiva em diversos trabalhos na literatura. Visando evitar este viés, foi utilizado um método apresentado por Damián (2017), Pang et al. (2006), Aldrich, An e Jansen (2019) que possibilita a avaliação da modelagem de tópicos por meio do cálculo da coerência de cada tópico. A etapa inicial na análise de coerência consiste em calcular para cada tópico o *Perplexity Mutual Information (PMI)*, também conhecido como similaridade vetorial (Equação 5.2). A etapa final na análise de coerência consiste por meio do cálculo da média das resultados do PMI. Equação 5.3, através da qual é obtido o valor da coerência de tópicos analisados $C(t)$.

$$PMI(w_i, w_j) = \log \frac{p(w_i, w_j)}{p(w_i) \times p(w_j)} \quad (5.2)$$

$$C(t) = \frac{1}{\sum_{i=1}^K \sum_{j=1, j \neq i}^K} \sum_{i=1}^K \sum_{j=1, j \neq i}^K PMI(w_i, w_j) \quad (5.3)$$

A coerência média de um modelo é dada a partir da pontuação de coerência em todos os tópicos (ALDRICH; AN; JANSEN, 2019). De acordo com (PANG et al., 2006), análise final normalmente envolve interpretações apenas nos tópicos mais coerentes de um modelo, em vez do modelo inteiro com todos os tópicos. Para isso, dado um modelo com um total de k tópicos, seleciona-se uma quantidade n (coerência) de tópicos de maior

coerência e coerência se a média destes valores. A fim de evitar o viés neste cálculo, o ideal é comparar as médias de diferentes valores de n e escolher o espaço de busca pelos tópicos mais coerentes dos modelos.

A identificação de tópicos nas reclamações foi realizada através do uso do algoritmo *Latent Dirichlet Allocation (LDA)* (LI et al., 2019). A implementação deste algoritmo foi realizada por meio da biblioteca *scikit-learn*² da linguagem Python.

O LDA é um modelo bayesiano utilizado por estudos relacionados na literatura (GENC NAYEN; AHRAN, 2017), sendo que diversos trabalhos demonstraram a eficácia do modelo na extração de tópicos em diferentes conjuntos de dados textuais (CHARLES-SMITH et al., 2015; LI et al., 2019; DANG et al., 2016). Na essência, o algoritmo assume que existem estruturas de objetos que não estão visíveis no conjunto de dados e usa a ocorrência de palavras observadas em diferentes instâncias (GÓMEZ et al., 2015). Dado um conjunto de dados textuais, a execução deste algoritmo retorna uma lista de tópicos associados a um conjunto de palavras. Este conjunto de palavras são associados ao tópico por meio de pesos e uma lista de registros (no caso deste trabalho, reclamações) com um vetor de valores de pesos, que definem a probabilidade de fazer um documento que contém um tópico específico (XIANG et al., 2017).

5.4 RESULTADOS

Foram coletados dados de empresas que atuam no setor de comércio e que tinham reclamações de consumidores nas duas plataformas. No total foram coletados, no período de 26/01/2019 a 26/01/2020, 289.666 reclamações nas duas plataformas. Na Figura 11 são apresentadas as taxas de reclamações proporcionais ao tamanho da população de cada estado. Sendo os estados da região sulista tem maiores taxas de reclamações por habitante.

Na Tabela 11 são expostos a quantidade de reclamações coletadas, juntamente com a taxa de resposta e de resolução fornecidas pelas plataformas no período de 6 meses. Ao observar a tabela, é possível constatar que o OGV tem menor quantidade de reclamações quando comparado com o RA, porém, proporcionalmente ao OGV há uma alta taxa de respostas de reclamações por parte das empresas analisadas.

As plataformas analisadas são similares em termos de perfis dos usuários e conteúdo dos textos. Entretanto, há particularidades em cada plataforma que se diferem a distribuição de reclamações e taxa de resposta. O RA é aberto para reclamação de todos as empresas disponíveis no mercado, qualquer cliente pode acessar a plataforma e enviar seu relato. Já o OGV só aceita reclamações de empresas que aderiram voluntariamente e se cadastraram

² <https://scikit-learn.org/stable/>

Tabela 14 - Número de reclamações por empresa em cada plataforma.

	Consumidor.gov			Reclamação		
	Reclamações	Respostas	Resolvidas	Reclamações	Respostas	Resolvidas
Americanas.com	10,456	94,5%	87,8%	23,357	90,2%	85,4%
Miguel's Luluz	5,004	96,7%	79,8%	10,400	90,2%	80,2%
Mercado Livre	38,346	100%	77,9%	71,048	8%	4%
Netshoes	4,377	99,8%	84,3%	47,337	8%	4%
Submarino	5,109	99,9%	75,2%	29,079	90,8%	84,2%

na plataforma. Isso tem impacto na quantidade de empresas presentes nas plataformas e é o principal motivo de haver reclamações sem resposta e resolução no RA. Tabela 14

As métricas textuais são pontos que destacam as diferenças entre as plataformas, tal como a credibilidade e tamanho das reclamações. Na Figura 10 são apresentadas as distribuições do tamanho das reclamações do CGOV e RA. A partir dos dados expostos, foi possível verificar que os relatos de problemas no CGOV (em azul) tem maior variação no tamanho dos relatos no RA (em vermelho), indicando a presença de duas bases de usuários com características distintas nas duas plataformas.

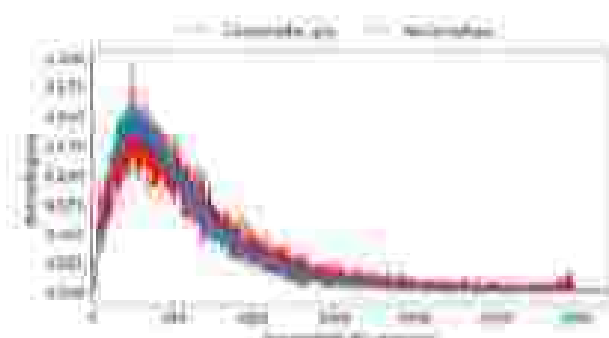


Figura 10 - Comparação do tamanho das reclamações nas duas plataformas.



Figura 11 - Mapa de distribuição das reclamações no Brasil.

Esta diferença fica mais evidente na Figura 12, que representa todas as reclamações coletadas das empresas das duas plataformas distribuídas em diferentes níveis de escolaridade. A segregação das reclamações em níveis de escolaridade nos revela um perfil mais subapoiado dos consumidores de cada plataforma de reclamações online. A seguir são descritas as características encontradas. No caso do CGOV identificamos os cinco principais grupos de pessoas cuja escolaridade é de 9^a a 10^a anos do ensino fundamental, seguido por "Ensino Médio e Nível Superior", "Ensino Acadêmico" e, por fim, pessoas com 1^a a 5^a anos do ensino fundamental. No caso do RA, encontramos um público de maior nível de escolaridade, sendo o principal grupo de pessoas identificar como que cursaram o "Ensino Médio e Nível Superior", seguido por ensino de nível "Ensino Acadêmico", 9^a a 10^a anos e 1^a a 5^a anos do ensino fundamental. Pode-se destacar que nas duas plataformas, o mesmo grupo de pessoas que realizam reclamações são as que tiveram pessoas acima do ensino formal, de 1^a a 5^a anos.

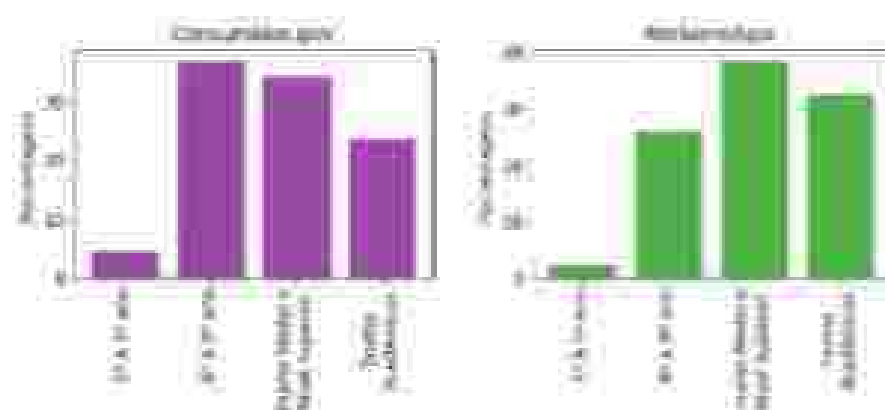


Figure 12 : Comparaison au niveau de compréhension entre un pilote et un expert.

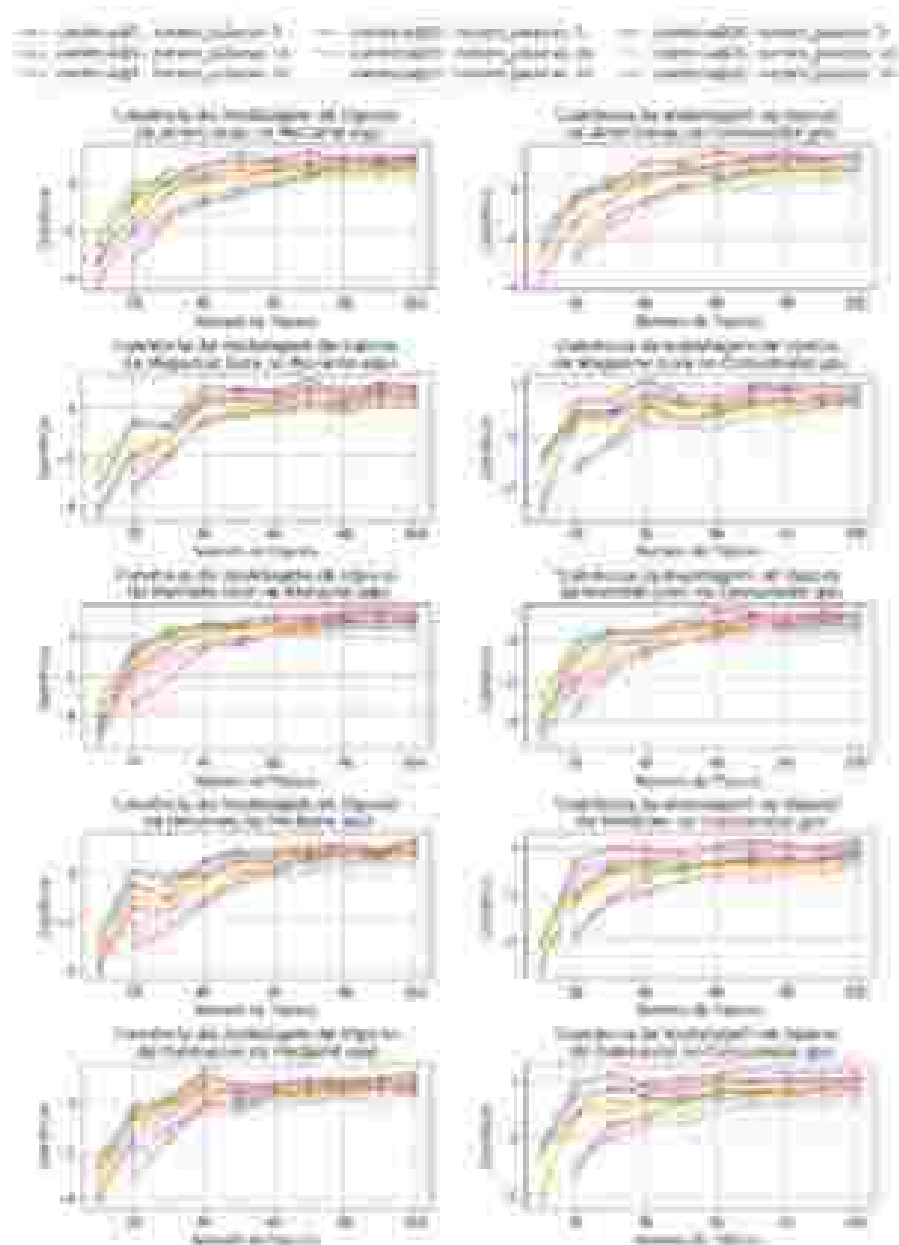


Figure 15 : Comparaison de la compréhension de l'expert

Os cálculos de coerência das empresas em diferentes questões são necessários para se ter uma modelagem de tópicos mais acurada e com maior nível de confiabilidade para determinar os principais tipos de problemas. Na Figura 13 são apresentadas os resultados dos cálculos de coerência para cada empresa a partir de diferentes questões que incluem: i) plataformas, ii) quantidade de tópicos, iii) número de palavras por tópico. Os maiores valores de coerência são considerados melhores e mais aplics a serem utilizados. Cada linha dos gráficos indica os resultados para um cenário diferente (i.e. "mercado105; numero_palavras5" indica que o cálculo foi realizado para os 5 principais tópicos que continham 5 palavras). A fim de determinar os melhores parâmetros a serem utilizados, os tópicos foram modelados considerando os seguintes valores de número de tópicos de [10, 20, 30, 40, 50, 60, 70, 80, 90, 100] variando a quantidade de palavras por tópico em [5, 10, 20].

Após final do processo, foi observado que houve variações nas quantidades ideais de tópicos modelados, tópicos principais e número de palavras para cada empresa e plataforma analisada. Na Tabela 15 são apresentadas os valores ideais para cada empresa, sendo: "Nº tópicos" - representa a quantidade de tópicos modelados pelo algoritmo; "Nº tópicos principais" - a quantidade de tópicos com maior coerência média que deverão ser utilizados nos resultados; "Nº palavras" - representa a quantidade de palavras para um tópico ter a melhor coerência.

Coerência pode ser observada sobre as empresas nos duas plataformas diferentes. Se com o número de tópicos e de palavras ideais para serem utilizados na modelagem, com exceção do "Submarino" que tem 10 como número de palavras ideais.

Tabela 15 - Melhores valores para modelagem de tópicos nas plataformas analisadas.

	Resultado Real			Resultado que		
	Nº tópicos Modelados	Nº tópicos Principais	Nº palavras	Nº tópicos Modelados	Nº tópicos Principais	Nº palavras
Amorcasma.com	70	5	5	60	5	5
Magazine Luiza	30	5	5	30	5	5
Mercado Livre	90	5	5	60	5	5
Netshoes	100	5	5	100	5	5
Submarino	40	5	10	20	5	5

A Tabela 16 apresenta os principais tópicos das reclamações nas duas plataformas, que representam os cinco principais tipos de problemas com maior valor e importância dentro de cada uma, considerando cada plataforma.

As células T1 a T5 de BA foram modeladas pois deram origem ao termo tópicos "Entrega". É importante ressaltar que cada tópico representa um elemento importante para a análise de reclamação, pois através deles pode o estabelecimento do contexto da empresa. De todas as células organizadas na tabela de termos a representar a importância através da ordem sequencial dos valores de coerência, sendo que na linha T1 estão os tópicos

Tabela 10 Tópicos modelados a partir das reclamações nos sites *Consumidor.gov* e *ReclameAqui*.

	Consumidor.gov				
	Americanas.com	Magazine Luiza	Mercado Livre	Netshoes	Submarino
T1	Cashback	Anuidade	Resolução de conflitos	Produtos	Produtos
T2	Acreditação	Expectativas	Publicidade	Entrega	Promoção
T3	Calibre	Entrega	União	Atendimento	Entrega
T4	Satisfação Insuficiência	Acreditação	Produtos	Clientes Insatisf	Entrega
T5	Preço	Debito	Serviço	Devolução	Cupom
	Reclame Aqui				
	Americanas.com	Magazine Luiza	Mercado Livre	Netshoes	Submarino
T1	Itens	Entrega	Resolução de conflitos	Entrega	Entrega
T2	Produtos	Clientes	Venda	Devolução	Devolução
T3	Entrega	Produtos	Cancelamento	Atendimento	Produtos
T4	Acreditação	Recurso pedido	Promoção	Clientes	Devolução
T5	Reembolso	Acreditação	Clientes	Exatidão	

mais representativos e na linha T5 estão os temas representativos dentro de principais. Os principais tipos de reclamações variam de acordo com a plataforma, e a seguir são exemplificados essas diferenças obtidas a partir dos tópicos mais relevantes.

- A empresa Americanas.com tem o tópico "Cashback" como problema principal no CGOV, o que indica que os usuários da plataforma têm mais problemas com esse método da empresa. Porém, no RA a principal reclamação é sobre a "Itens", que pode estar relacionada com os mais variados serviços da empresa, o que significa que é necessário investigar qual é como os serviços lidam com a base dos consumidores.
- A empresa Magazine Luiza tem o tópico "Anuidade" como problema principal no CGOV, onde o grupo de clientes que são usuários do CGOV também a ter mais problemas com os produtos com anuidade. Para o grupo de usuários do RA a principal reclamação é sobre a "Entrega", o devido as características da empresa essas reclamações estão extremamente relacionadas a entrega de produtos.
- A empresa Mercado Livre tem o tópico "Resolução de Conflitos" como problema principal no CGOV e no RA, o que indica que independentemente do grupo de cliente que utilizam as plataformas a principal reclamação é a mesma, indicando que o problema é recorrente e afeta muitos clientes.
- A empresa Netshoes tem o "Produtos" como problema principal no CGOV e "Entrega" como principal reclamação no RA. Estes tipos de reclamação são diferentes e devem ser tratados individualmente, mas devido ao modelo de negócio da empresa estão muito relacionados entre si.

- A empresa Submarino tem o “Produto” como problema principal no OGIV e “Entrega” como principal tipo de reclamação no RA. Porém, devido ao modelo de negócio da empresa, esses tipos de reclamações não necessariamente estão relacionados entre si.

Os achados aqui apresentados se mostram valiosos e com grande possibilidade de orientar a atuação de empresas nas duas plataformas. Através das respostas, é possível obter uma ordem sequencial de problemas recorrentes que desorientam a relação entre empresa e consumidor. Além disso, devido a presença de grupos distintos de usuários nas plataformas, é possível estruturar estratégias de solução que se adaptem a cada grupo.

5.5. CONSIDERAÇÕES FINAIS

Neste artigo, foram analisadas reclamações de cinco empresas do setor de e-commerce brasileiras em duas plataformas de reclamações online. O objetivo destas análises é fornecer uma base para a compreensão das oportunidades que elas oferecem para orientar a tomada de decisões em negócios. Foram aplicadas técnicas de mineração de dados visando a fim de explorar as plataformas e extrair suas características. Os resultados mostram que apesar de as plataformas terem a mesma finalidade, elas divergem em grupos de usuários, confiabilidade de solução, conteúdo e principais tópicos das reclamações. Ademais, as análises se mostraram viáveis mesmo com grandes quantidades de dados, o que torna possível a antecipação e correção de problemas de forma rápida e eficaz, e abre a possibilidade de fazer análises globais do mercado e convergência de acordo com a perspectiva de cada plataforma.

Diante disso, a principal contribuição deste estudo está na produção de uma abordagem de análise de reclamações para as duas plataformas que permite identificar diferentes formas de atuação que uma empresa deve seguir para melhorar o relacionamento com seus clientes. No entanto, não foram realizadas análises de significância estatística nas comparações apresentadas neste estudo. E as análises e resultados obtidos são apenas aplicações preliminares nas empresas. Como trabalhos futuros, pretende-se expandir as análises de modo a incluir análise de significância estatística nas comparações, trabalhar com outros problemas em empresas que explore tanto as reclamações quanto a respostas.

5.6. AGRADECIMENTOS

Os autores agradecem ao Serviço de Interdições Abaisdo (DAAD), à Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico de Matochdo (FAPESMA) e a Universidade Federal do Oeste do Pará (UFOPA) por meio da Pró-Reitoria da Cultura, Comunidade e Extensão (PRCCF), pelas ferramentas disponibilizadas à execução desta pesquisa.

6 ARTIGO - GAINING INSIGHTS ON STUDENT SATISFACTION BY APPLYING SOCIAL CRM TECHNIQUES FOR HIGHER EDUCATION INSTITUTIONS

ABSTRACT

Social Media and Customer Relationship Management (CRM) are already widely used in business settings, but other non-commercial sectors started only recently to adopt them. Among them are Higher Education Institutions (HEIs). Even though research shows positive effects on the quality of services, student satisfaction, and attractiveness towards international students, the adoption is very low. This research in progress reviews the state of research about Social CRM in HEIs and gives an example of the potential of social media for CRM approaches of HEIs by applying Social CRM concepts and techniques for better understanding the negative service experiences of students. By applying analytical Social CRM techniques on large amounts of User-Generated Content (UGC) in comments platforms the paper gives insights into problems change inaccessible with manual methods. Based on the current research about Social CRM as well as the demonstrated potential of social media for CRM strategies of HEIs, this paper concludes with a call for further research on Social CRM in HEIs.

Keywords: CRM, Social Media, Student Satisfaction, Complaint Management, Text Mining, Topic Modeling

6.1 INTRODUCTION

Social media gain importance in higher education not only since the COVID-19 pandemic prevents personal contact between students and professors. Researchers investigate the use and potential of social media in specific areas for several years. For example, as support in lectures (DYSON et al., 2015), support in hybrid learning environments (LI, 2011), for marketing purposes (KARNA, SUPRIANA, MAULIHVI, 2015), or examined how students use them for study purposes (BRASTINSKI, ATRAEF, 2012). However, their potential for building relationships with students and managing the student life cycle was only sparsely examined. A reason may be the fact that many universities do not compete over the increasing number of students or that the relationship between students and professors is more focused on teaching and students are expected to actively manage their study program. But as an increasing amount of young people, and so students, use social media more frequently, they also use them during their student lifecycle and expect higher education institutions (HEIs) to do the same.

Applying concepts and techniques developed for the management of customer relationship management (CRM) with the help of social media could help to manage this transformation. As in the industry, wherein many cases until 2010 the enterprise still missed the customer experience (GREENBERG, 2010), the student experience is still often missed by HEIs. Universities already pay attention to social media to maintain communication (DYSON et al., 2012) but use them only for dedicated tasks. Greenberg (2010) pointed out, that enterprises need first to figure out the business models, applications, processes, and social characteristics that are required to actually implement the social CRM before social media customer service begins to happen. This counts now for HEIs as they need to figure out the application areas of social media within the student life cycle and service offerings of the university.

This paper aims to initially explore the relevance of social media by analyzing the use of such channels in critical steps within the student life cycle, namely the handling of complaints as part of the service phase. As visible in a famous example of Social CRM (JARVIS, 2005), students will use social media if they are not satisfied with the service experience, regardless of whether HEIs are active on social media or not. Following the concept of Social CRM (GREENBERG, 2010, ALT, REINHOLD, 2010), higher education could build up its presence on social media platforms, provide services by using social media as a channel to workloads, learn from the content in social media and use these channels to perform collaborative tasks with students. Actively using social media and providing a satisfying service experience may decrease the number of complaints. As research about the management of complaints by HEIs via social media is scarce, this paper aims to show that students use social media for complaints, that a link between the number of complaints in social media with the active provision and management of

social media by universities exists and to identify the core topics of student complaints. The research questions are:

- [RQ1] Are students using external and public platforms to complain about education-related services along their student life cycle?
- [RQ2] How can we derive information about major service quality issues affecting student satisfaction?
- [RQ3] What types of insights on customer satisfaction can HEIs managers expect from an analysis of external complaints?

The remainder of the paper is structured as follows. First, research about the role of customer satisfaction and Social CRM is reviewed and key elements for assessing customer satisfaction with the help of Social CRM are identified. Second, an experiment demonstrates an approach for analyzing customer satisfaction in higher education. Third, the results from the experiment are discussed, answering the research questions.

6.2 CRM AND SOCIAL CRM IN HIGHER EDUCATION

In this section, we discuss the CRM and Social CRM applied in high education institutions to improve the services & the students' satisfaction. We first outline the effects of the CRM on the service quality and students' satisfaction in HEIs, and before we discuss the potential of Social CRM to understanding customer satisfaction and to managing the service quality.

6.2.1 CRM AFFECTS SERVICE QUALITY AND STUDENT SATISFACTION IN HEIs

The application of CRM concepts in higher education are examined from different perspectives already. [Raj et al. \(2016\)](#) show that the main principles of CRM can also be applied to the context of HEIs and that HEIs must consider more stakeholders than just students in their CRM approach, calling for also using social media for linking and interacting with numerous stakeholders. [Nair, Chao & Yang \(2007\)](#) point out that HEIs first need to understand the student lifecycle (Student → Prospect → Applicant → Admitted → Enrolled → Alumni) before successfully making use of a CRM approach.

By analyzing social media content, HEIs can improve their understanding of the student life cycle and optimize their CRM. [Rajeev \(2015\)](#) shows that student satisfaction is a good indicator for the successful adoption of CRM for the creation of a student-oriented environment and constantly adapting its processes. Critical elements are the university organization and management of teaching processes, academic staff skills and competencies, management board activities and institutional development, and quality of study materials

used in the classes, and application of learning methods. [Seriman & Oflaz \(2016\)](#) points out that universities that aim to achieve a leadership position in the higher education sector need to show an additional focus on reproducing highly skilled faculty staff that will have a capacity to improve teaching with regard to technology changes and market requirements. It is also important that the university board and leading people such as deans of HEIs understand the strategic dimension of CRM orientation at universities. Both call for HEIs to adopt social media early as new technology and to develop an integrated management approach for social media and CRM. A study from [Seriman & Yildiz \(2018\)](#) shows the handling of students as customers provides a competitive advantage for higher education and enhances a college's ability to attract, retain and serve its customers. The benefits of implementing CRM in a college setting include a student-centric focus, improved customer data and process management, and increased student loyalty, retention, and satisfaction with the college's programs and services. As colleges increasingly embrace distance learning and e-business, CRM will become more pervasive. The COVID pandemic further increased this need. [Hachimi et al. \(2017\)](#) confirm this observation by showing that implementing electronic CRM can cause customer satisfaction, loyalty, retention, and high service quality as students pointed to be a customer.

CRM supports the understanding of customer expectations and thus provides a basis for service customization, which in turn can positively affect service quality. [Wah & Wright \(2016\)](#) show that an effective CRM program to improve service quality affects customer satisfaction and even has the ability to induce positive advocacy behavior from the international students. A key element is gaining and understanding customer's experience feedback.

6.2.2 NEW POTENTIALS FOR UNDERSTANDING CUSTOMER SATISFACTION AND MANAGING THE SERVICE QUALITY ARISE FROM SOCIAL CRM

A key element of Social CRM is the interaction with stakeholders that influence the service system of a business and the knowledge derived from this interaction. As [Greenberg \(2011\)](#) points out, the customer becomes the focal point of the ecosystem, and service providers can make use of social media to understand their role in the customer ecosystems. Social CRM provides the means for that, but as [Soliman et al. \(2017\)](#) a social CRM model for HEIs is virtually non-existent. However, the first research points out, that applications, data, and information, adapted business processes, social media presence are among the critical success factors for social CRM in HEIs.

Following [Meyyappa, Hidayatullo & Ratuandjo \(2015\)](#) many universities started with web 2.0 and social media adoption has focus mainly on real-time events website, widgets, and social networks for the users of the university website. The study of [Chen et al. \(2015\)](#) shows that Social CRM propels HEI to engage in dialogical conversations and collaborative

relationships. The use of social media platforms, allowing to reshape the HEI-student formal relationship, strengthening educational bonds through the development of dialog, provides mutually beneficial value and, ultimately, allows for the growth of social and educational communities.

However, only a few examples have further investigated the potential of analytical Social CRM for HEIs, especially for assessing student satisfaction with experience service quality. [Bardhanja et al. \(2017\)](#) show in general by mapping the student lifecycle with CRM processes and features of Social CRM that the latter can support core CRM processes of HEIs, and the application of Social CRM software can support operational, analytical, and collaboration tasks. [Karia, Srijana & Mondal \(2015\)](#) show how data analysis can provide further insights on candidates and help to individualise marketing activities. But unlike Social CRM in a business context, the potential of monitoring and mining for HEIs has not been understanding.

6.3 IMPROVING THE UNDERSTANDING OF NEGATIVE SERVICE EXPERIENCES IN HEIS WITH ANALYTICAL SOCIAL CRM TECHNIQUES

This section discusses the use of analytical Social CRM to understand the students' negative experiences in HEIs on Social Media platforms. In the following subsections, we present an overview of the process to analyse from social media platforms, the methodology to perform the analysis, the potential data source, and the potential analysis methods.

6.3.1 COMPLAINT AND SATISFACTION ANALYSIS IN EXTENSIONAL SOCIAL MEDIA

Following the concept of analytical Social CRM ([JOURNAL OF AUF, 2011](#)), building up a customer feedback and satisfaction analysis requires scanning the data of social media platforms where students share feedback and opinions. In a second step, this data needs to be turned into knowledge about customer satisfaction by either manual evaluation, observation of keywords, or applying methods for understanding at the semantic level. While the manual evaluation can provide rich insights, the potential amount of data makes it challenging for HEIs. Applying basic analyses such as looking for trending topics and sentiments of students might support HEIs to understand the behavior of their students. Although these analyses are supported by many social media monitoring tools, they have limits in uncovering larger patterns, such as the relationship between related issues. The understanding previously unknown factors that affect service quality methods of data mining need to be applied.

6.1.2 PROCESS DESIGN

In this research, it is critical to map relevant data sources and evaluate all pertinent analyses that improve customer satisfaction. One well-known data analysis methodology for conducting real-world projects is the Cross Industry Standard Process for Data Mining (CRISP-DM). This methodology proposes a comprehensive process model for carrying out data mining projects. The process is divided into phases of *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment* and is independent of the industry sector and the technology used (WHITE, 2006).

In this research, CRISP-DM will be applied for the analysis of student complaints in Brazil. HEIs in Brazil provides a good example, because of the availability of well-used independent social media platforms for publicly raising and discussing complaints independently from specific industry sectors. Thus, insights on service quality and customer satisfaction outside of a HEIs direct control are accessible.

6.1.3 POTENTIAL DATA SOURCES

The second phase of CRISP-DM is called “Data Understanding”, in which it is possible to identify issues on data and provide initial insights into available data. However, there is plenty of platforms with a large volume of User Generated Content (UGC) available with some challenges such as data diversity, unstructured data, missing data, etc (LOBATO et al., 2017). Despite those challenges, such platforms represent an interesting data source for providing business insights with reduced costs when compared with customer surveys and other market research strategies (TAUJAR, MUDA, 2016; VERMEER et al., 2019). In the context of Brazilian HEIs relevant data sources are for example:

- *Consumidor.gov* (CONSUMIDOR.GOV, 2021) - Contains data referring to complaints reports about universities. The data types include strings (raw text), timestamps (Date and Time), and numerical features. This platform does not provide an *Application Programming Interface* (API) for data acquisition. However, considering that it is a public platform with open data (considering Brazilian Legislation), the data can be requested using the national accountability platform.
- *Ministry of Justice and Public Security (MJSP)* (MJSP, 2021) - The MJSP portal provides details about the complaints published on the *Consumidor.gov* platform. Large parts of this data are in textual format. This platform provides an interface to get the available data.
- *Data from Higher Education Census (DECT)* (DECT, 2021) - It is data related to the 2019 higher education census, and contains information about Brazilian students, courses, and high educational institutions. Most of the data are numerical features, requiring a mapping procedure with a data dictionary.

- **Data from Social Media platforms** - Contains information related to followers, publications, and Comments. However, it is important to verify the privacy rules and API restrictions of each platform.

6.3.4 POTENTIAL METHODS FOR ANALYSIS

"Data preparation" and "Modeling" phases of the CHES-DM methodology represent the core for Computer Scientists (WIRTH, 2000). The pre-processing methods aim to improve data quality, consequently, improving the reliability of the results (LOHATO et al., 2017). Considering that most UCLs are created in text, there is a lot of effort on the development of text mining methods and pipelines (FERNANDES et al., 2018). The canonical pipeline includes the application of specific pre-processing methods for text data (ALLAHYARI et al., 2017; GARCIA et al., 2017; HAN et al., 2016). It can be followed by a data fusion/enrichment step, aiming to combine different data sources, which can reduce bias and uncertainties, increase reliability, and improve accuracy (ZHAO, 2011; QI et al., 2020).

The data preparation is followed by the data analysis itself. There are common tools in Text Mining. For instance, topic modelling allows the identification of the most frequent topics and their terms, which would be difficult to discover through a manual process that (RODRIGUES et al., 2011; PAUL, CHAN, KWON, 2006). The topics discovered that can be modeled as a graph, indicating the relationship between the topics, and allowing the identification of terms clusters (BRADLEY, KLEINBERG, 2010). In addition to that, sentiment analysis can be used to extract the feeling through automatic polarity detection (JO, PEDRUELA, 2017; RAVI, RAVI, 2015), and the text quality can be assessed by the evaluation of the legibility and by the determination of the quality according to the expected in each phase of regular education (THERIAULT et al., 2017; THERIAULT, 1988; GUINAN et al., 2012).

6.4 DEMONSTRATION

This section presents an analysis of the textual content of students' complaints about two higher education institutions in Brazil, named *University A* and *University B*. These two universities have multiple internal channels of customer services available to students, which include the website online chats, email, telephone and social networks (see Table 17). On social networks, these universities have in general a high number of followers. However, *University A* has a low number of interactions on their content, opposite to *University B* with a very number of interactions. Furthermore, their students are using external platforms for making complaints. As it is known that on average two thirds of consumers check product, service, and brand evaluations before deciding on the purchase, (BE et al., 2018; CONSTANTINIDES, HOLLESCHOVSKY, 2016) these external complaints can

have an impact on reputation and in consequence the willingness of future students to join the university.

	University A	University B
Total of Students (May 7, 2021)	30,311	30,776
Digital Device Channels	- Website - Phone - WhatsApp - Social Networks: - Twitter with 4,994 followers - Instagram with 62,880 followers - Facebook with 611,780 followers - YouTube App - Consumidor.gov	- Website - Phone - WhatsApp - Social Networks: - Instagram with 261,080 followers - Facebook with 1,329,863 followers - YouTube App - Consumidor.gov
Complaints on Consumidor.gov ⁶	- Total of Complaints: 1,100 - Solution Rate: 63.1%	- Total of Complaints: 1,060 - Solution Rate: 57.4%
Complaints on YouTube App ⁷	- Total of Complaints: 11,394 - Solution Rate: 34.8%	- Total of Complaints: 12,226 - Solution Rate: 32%

Table 17: Summary information about the universities

Following the methodology described before, the analysis comprised a pre-processing step and the analysis phase, using topic modeling and topic correlation methods. The data were collected on Consumidor.gov⁸ (CONSUMIDOR.GOV, 2021) from January to March of 2021, with a total of 452 complaints about University A and 810 about University B. Topic modeling was carried out using Negative Matrix Factorization (NMF) (JEDN et al., 2019) and revealed the ten main topics in complaints about each university (see Figure 14). Some topics are not exclusive, indicating that the two institutions facing the same type of problem, related to the "Payments", "Attestation of degree", and "Classroom".

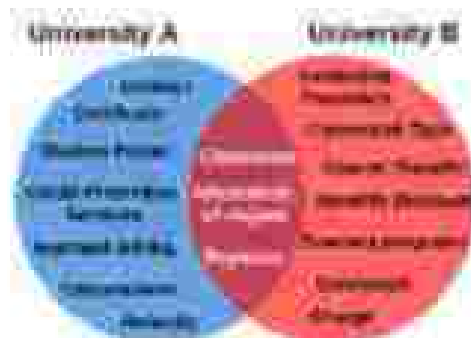


Figure 14: Main Topics in complaints of the University A and University B

The correlation of the topics allows the identification of relationship degrees between the topics. Figure 15 was produced using open-source software called Gephi⁹, and it shows the relationships between most frequent topics/objects, and the strongest blue tone of the

⁶ *Disclaimer:* It is important to emphasize that all data collected on Consumidor.gov will be used exclusively for proof-of-concept purposes. The authors have no interest in the commercial use of this data.

⁷ <http://gephi.org/>

line indicates a stronger degree of correlation. In the topics of *University A* (see Figure 17a) the formation of a chain of main problems refer to Payments, Refunds, and the Contact. Figure 17b shows the chain of main problems for *University B*, which are related to Payments, Charges, Monthly Discount, and Enrollment.

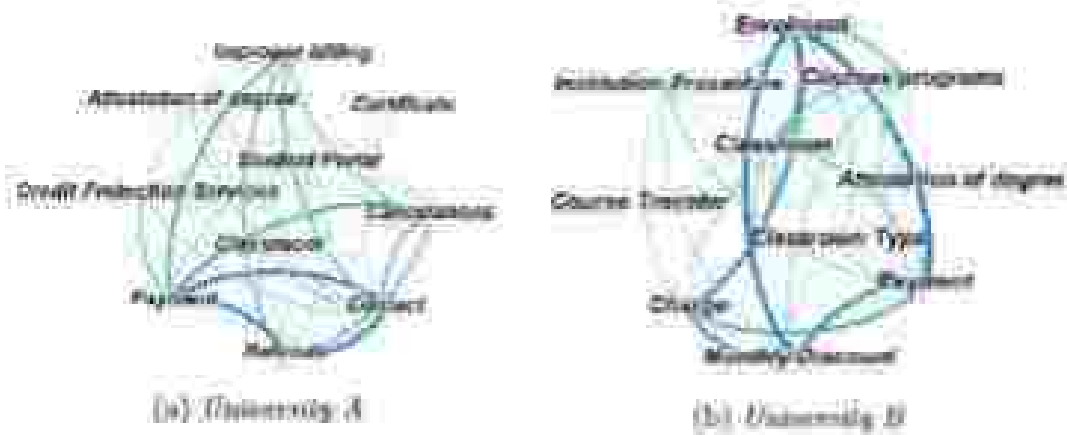


Figure 17 Topic correlations of complaints

Modeling and correlation of topics give HEIs managers insights on potential service quality issues. But, they are also a starting point for planning actions to solve the problem chains that make up principal causes of dissatisfaction among students. For example, among the problems of *University A* are several types of issues related to finance and customer service. They indicate students are dissatisfied because of financial issues together with the university customer service. For solving these problems, more detailed analysis can be made on complaints related to these areas, to understand and identify products and services that are the source of the problems.

The demonstrated analysis provides insights based on a large number of complaints related to two universities. Such aggregated insights on major students' problems and their relations can support decision-making and as a basis for creating efficient actions for solving the chain of issues at the root of student complaints. Thus they can contribute to CRM efforts by helping and relating to developing better products and services centered on the expectations and needs of students.

6.5 CONCLUSION AND IMPLICATIONS

This paper investigates the potential of Social Media for improving service quality and student satisfaction as part of the ever-increasing adoption of CRM in HEIs. Building upon recent research about critical success factors of CRM for HEIs, the paper demonstrates how universities can improve their service systems with insights from Social CRM analysis. Based on an examination of selected universities in Brazil this research shows that students use external platforms to complain about their service experiences, regardless if

the university provides its own platforms (RQ3). With the application of data analysis and mining techniques employed in social CRM (FERNANDES *et al.*, 2020; FOLDSWELL; ALT, 2011) in the enterprise context, major issues can be efficiently summarized (RQ2) from large data volumes. HEIs can learn from this data (RQ3) about the service quality experience of students and factors that negatively affect student satisfaction. It is also a basis for comparison and evaluation with other HEIs and an opportunity to identify more successful approaches. In addition, from the results produced by the text mining tools used in the demonstration, it was possible to observe that complaints are a rich source to extract knowledge about students and services offered by universities.

Besides these direct insights, this research provides, on a more general level, a further example of the benefits of Social CRM for HEIs and the importance of further research about the application fields and implementation approaches for them. Only an integrated Social CRM allows to get data from internal and external sources and to use obtained knowledge for improving services. Many insights on the successful integration of Social Media with CRM in enterprises are available and HEIs can build on this extensive knowledge while adapting it to their specific environment.

This paper is an example of the potential of advanced semantic analysis in a limited case. Therefore, the paper has limitations. The obtained insight is not generalizable and demonstrated only the insights that can be acquired by using the proposed approach. Also, the insights have not been evaluated with managers in HEIs in terms of novelty and meaningfulness compared to other sources. Therefore, this research is planned to be replicated in the next step, by using the approach for a larger set of universities, expanding the insights by comparing the results in a larger set of universities and evaluating results of managers of HEIs regarding their contribution towards a better understanding of student satisfaction and potential improvements in service quality.

7 CONSIDERAÇÕES FINAIS

Esta dissertação foi composta por um compêndio de estudos relacionados à análise inteligente de mídias sociais para potencializar a gestão de relacionamento com clientes. Diferentes abordagens de inteligência foram estudadas e testadas neste trabalho com o objetivo de melhorar os sistemas de Social CRM. Os quatro artigos apresentados demonstraram que os objetivos propostos do projeto foi atingido.

Além dos estudos apresentados nos capítulos 3, 4, 5 e 6, outros trabalhos foram publicados, os quais representam contribuições marginais a este trabalho e foram incluídas nos Apêndices A e B. A seguir, cada um desses artigos é brevemente descrito:

1. Artigo: *“Instrumentos para Análise de Mídias Sociais: Um levantamento sistemático”*

- Este estudo teve como objetivo principal inspecionar o estado da arte e o estado da prática em relação à análise de mídias sociais, para identificar as bases, métodos e ferramentas mais utilizadas pelos pesquisadores em suas análises. Os resultados deste estudo têm potencial de orientar futuras pesquisas, pois poderão ser corroboradas de maneira mais objetiva, de modo a economizar tempo e esforço nas fases de análise de fontes de dados, além de apontar para as pesquisas na tecnologia de análise de dados mais utilizadas. O manuscrito completo deste artigo está disponível no Apêndice A desta dissertação.

2. Artigo: *“Social CRM as a business strategy: developing dynamic capabilities of Micro and Small Enterprises”*

- Este estudo teve por objetivo principal estudar o uso de Social CRM por Micro e pequenas empresas, a fim de identificar oportunidades de intervenção. A pesquisa deste estudo ocorreu por meio da aplicação de um questionário em 31 empresas e análise de estudos de caso conduzidos em quatro empresas. Os resultados permitiram a identificação de pontos críticos de falha na utilização de Social CRM por micro e pequenas empresas, e que a utilização de Social CRM ainda é pouco explorada por essas empresas. O manuscrito completo deste artigo está disponível no Apêndice B desta dissertação.

De acordo com os resultados deste estudo, acredita-se que as abordagens propostas e estudadas nos artigos contribuirão para a melhoria de sistemas de Social CRM por meio:

- Da automatização da avaliação da efetividade de comunicações empresariais. Pois eliminam o risco da avaliação manual e permite a análise de grandes quantidades de dados de forma eficiente e rápida;

- A partir da análise do potencial, da identificação das oportunidades e desafios e da demonstração da aplicabilidade em um cenário real, foi aprovada e validada a viabilidade do uso de tecnologias online para o aprimoramento de sistemas de Social CRM;
- Do auxílio na escolha da base de dados e métodos para se analisar. O que otimiza o início do processo de extração de conhecimento das mídias sociais;
- Do entendimento dos principais problemas enfrentados por micro e pequenas empresas na utilização do Social CRM. O que permite a implementação personalizada dos sistemas de Social CRM por esse setor da economia.

Além do impacto científico demonstrado pela publicação dos seis artigos representados, o presente trabalho também contribui para com o estado da técnica, fornecendo materiais e métodos para a condução de análises de dados de mídias sociais por outros pesquisadores, por meio da disponibilização de uma base de dados de treinamento para auxiliar na classificação de conteúdos empresariais e de uma pipeline de análise que inclui diversos métodos e dá início o processo de extração de conhecimento das mídias sociais. A base de dados foi tornada pública⁴ (SOUZA, JUNIOR, LOBATO, 2021) e a pipeline será submetida ao registro de softwares junto ao Instituto Nacional da Propriedade Industrial (INPI).

7.1 TRABALHOS FUTUROS

Como trabalhos futuros, planeja-se a adoção e aprimoramento dos métodos presentes na pipeline de análise. Para isso, serão estudadas e avaliadas as possibilidades de melhorar o processamento da linguagem natural através do uso de *reinforcement learning*, *word embeddings*, *deep-learning* e algoritmos como *Directional Encoder Representations from Transformers* (BERT), *Convolutional Neural Network* (CNN), *Hierarchical Text Classification* (HTC), dentre outros.

Planeja-se também a construção da documentação da pipeline de análise de dados. Isso permitirá que a manutenção e o uso da ferramenta sejam facilitados, uma vez que todos os métodos e formas de uso foram descritos e exemplificados. Além disso, essa documentação permitirá que novos membros sejam integrados à equipe de desenvolvimento de forma simples e rápida, facilitando a ampliação da equipe de testes e análises disponíveis na pipeline.

Por fim, planeja-se realizar uma análise de viabilidade de mercado para implementação da pipeline. Este processo permitirá o mapeamento e avaliação do potencial econômico, bem como a identificação de oportunidades e desafios para a criação de um

⁴ <https://doi.org/10.2201/revista.344286>

produtos a partir do pipeline construído neste trabalho. Desta forma, pode-se elaborar um plano de negócios e implementar um protótipo de produto considerando as reais necessidades e desafios impostos pelo mercado.

7.2 DIFICULDADES ENCONTRADAS

Durante o desenvolvimento deste projeto, algumas dificuldades e dificuldades foram encontradas. No início dos estudos foram utilizadas ferramentas implementadas com Java¹ e com a biblioteca Scikit² do Python, porém observou-se que essas ferramentas não eram eficientes e demandavam muito tempo para coletar e organizar os dados, com o avanço dos estudos, as dificuldades técnicas foram superadas e novas ferramentas de coleta foram construídas a partir da biblioteca Requests³ do Python, o que tornou o processo mais rápido e otimizado. Além disso, a conexão com a internet foi um problema constante na coleta inicial de dados, diversas soluções foram analisadas e testadas, como a utilização da plataforma Kaggle⁴, que permitia a execução por 6 horas, e normalmente, o processo era interrompido antes do fim. Assim, foi desenvolvida a construção de um servidor virtual privado com acesso ininterrupto à internet para a execução das ferramentas de coleta de dados necessárias às análises.

Após a coleta de dados, a execução das análises foi dificultada pelas rotinas computacionais disponíveis. Esta limitação teve um grande impacto nas tarefas de pré-processamento, devido ao tamanho do conjunto de dados, mas processos exigiram muitas recursos computacionais e demandaram horas para serem concluídas, de acordo, era necessário dividir os dados em conjuntos menores para que poderiam ser executados dentro das 6 horas fornecidas pela plataforma Kaggle. Da mesma forma, as tarefas de avaliar o modelo através de técnicas para validação foram impactadas, mas neste processo, não foi possível dividir o banco de dados em conjuntos menores, pois a avaliação deveria ocorrer considerando todos os dados, portanto esse processo levou em execução por vários dias até a sua conclusão.

Além destas, encontraram dificuldades na validação dos resultados obtidos nas análises de dados. Este estudo envolveu a análise de dados de várias empresas de diferentes setores, mas não foi possível estabelecer contato com elas para que os resultados fossem validados e utilizados internamente. Além disso, não foi possível estabelecer contato direto com especialistas em Social CRM que atuam no Brasil, e os resultados deveriam ser avaliados e validados por especialistas que atuam no exterior, e devido a essa limitação com o domínio de aplicação e com o contexto brasileiro, este processo foi realizado em diversos locais, que incluíram a contextualização e implementação das análises.

¹ <https://developer.mozilla.org/pt-BR/docs/Web/JavaScript>

² <https://scikit-learn.pythonsciencetools.com/>

³ <https://docs.python-requests.org/en/latest/>

⁴ <https://www.kaggle.com/>

REFERÊNCIAS

- ABCCOMM. Crescimento do e-commerce no Brasil. Disponível em: <http://abccomm.org/noticias/crescimento-do-e-commerce-no-brasil>. Acesso em: 26/03/2021. 2019. Disponível em: <<https://abccomm.org/noticias/crescimento-do-e-commerce-no-brasil/>>. Citado na página 49.
- AGHASIAN, E.; GARG, S.; MONTGOMERY, J. An automated model to score the privacy of unstructured information – Social media case. *Computers and Security*, Elsevier Ltd, v. 92, p. 101778, 2020. ISSN 0167-1078. Disponível em: <<https://doi.org/10.1016/j.cose.2020.101778>>. Citado na página 25.
- Agência Nacional de Telecomunicações (ANATEL). Anatel – Paraná. 2020. Disponível em: <<https://www.anatel.gov.br/paises/ocmco/paísesparana>>. Citado 2 vezes nas páginas 24 e 52.
- AHMAD, S. N.; LAROCHE, M. Analyzing electronic word of mouth: A social commerce construct. *International Journal of Information Management*, Elsevier Ltd, v. 37, n. 3, p. 202-213, 2017. ISSN 1058-0112. Disponível em: <<https://doi.org/10.1016/j.ijinfmangt.2016.08.001>>. Citado 4 vezes nas páginas 25, 32, 39 e 51.
- ALDOUS, K. K.; AN, J.; JANDON, D. J. View , Like , Comment , Post : Analyzing User Engagement by Topic at 4 Levels across 2 Social Media Platforms for 53 News Organizations. In: *Proceedings of the Thirtieth International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2019. p. 47-57. Citado 4 vezes nas páginas 33, 47, 48 e 70.
- ALEXA. Alexa - Top Sites in Brazil – Alexa. 2019. Citado 2 vezes nas páginas 54 e 55.
- ALI, M.; HOORABCHI, M. E.; MESBAIE, A. Same Apps, Different App Stores: A Comparative Study. *Proceedings - 2017 IEEE/ACM 10th International Conference on Mobile Software Engineering and Systems, MOBILESoft 2017*, p. 79-90, 2017. Citado na página 11.
- ALLAHYARU, M. et al. A Brief Survey of Text Mining, Classification, Clustering and Extraction Techniques. 2017. Citado 3 vezes nas páginas 28, 29 e 30.
- ALMEIDA, G. R. de; CIRQUEIRA, D. R.; LOBATO, F. M. Improving social CRM through electronic word-of-mouth: a case study of reclusosapi. In: *SIBIC: Anais Eletrônicos do XVIII Simpósio Brasileiro de Sistemas Mult mídia e Web [S.l.]*, 2017. p. 107-110. Citado 2 vezes nas páginas 51 e 53.
- ALMEIDA, G. R. E. de; LOBATO, F.; CIRQUEIRA, D. Improving Social CRM through electronic word-of-mouth: a case study of ReclusosApi. *XIV Workshop de Trabalho de Indução Científica*, 2017. Citado 2 vezes nas páginas 21 e 23.
- ALMEIDA, J. S. DE N. G. R. T. de; LOBATO, F. M. F.; JUNGION, A. F. L. J. Melhorando Sistemas de Social CRM por meio de Eletronic Word-of-Mouth. *Revista Eletrônica de Indução Científica em Computação*, v. 17, n. 4, 2018. ISSN 1549-8219. Citado na página 67.

- ALT, R.; REINHOLD, O. Social Customer Relationship Management (Social CRM). *Business & Information Systems Engineering*, v. 4, n. 3, p. 287-291, 2012. ISSN 1867-0212. Disponível em: <<http://link.springer.com/10.1007/s12599-012-0225-5>>. Citado 2 vezes nas páginas 23 e 31.
- ALT, R.; REINHOLD, O. Social crm: Challenges and perspectives. In: *Social Customer Relationship Management* [S.l.]: Springer, 2020. p. 61-102. Citado na página 23.
- ANAYEL, *Model - Personal - Files - condicional*, 2020. Disponível em: <<https://condicional.gov.br/index.php>>. Citado 2 vezes nas páginas 11 e 62.
- BADWAN, J. J. et al. Adopting technology for customer relationship management in higher educational institutions. *IAHW*, 2017. Citado na página 80.
- BAEFAR, A. Z.; MUDA, M. The Impact of User-Generated Content (UGC) on Product Reviews towards Online Purchasing - A Conceptual Framework. *Procedia Economics and Finance*, v. 37, p. 337-342, 2016. ISSN 2212-6671. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2212667116303341>>. Citado 3 vezes nas páginas 51, 57 e 62.
- BANKOLE, F. O.; OSEI-IBIVSON, K. M.; TOWERS, I. The Impacts of Telecommunications Infrastructure and Institutional Quality on Trade Efficiency in Africa. *Information Technology for Development*, v. 31, n. 1, p. 29-43, 2015. ISSN 15740176. Citado na página 74.
- BARATA, G. M. et al. Social CRM in Digital Marketing Agencies: An Extensive Classification of Services. 2018 *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, IEEE, p. 770-783, 2018. Citado na página 67.
- BARRITO, A. M. The word-of-mouth phenomenon in the social media era. *International Journal of Market Research*, SAGE Publications Sage UK: London, England, v. 56, n. 3, p. 633-674, 2014. Citado na página 70.
- BELLACHIORE, G.; JUNG, J. J.; CAMACHO, D. Social big data: Recent achievements and new challenges. *Information Fusion*, Elsevier B.V., v. 28, p. 45-59, 2016. ISSN 1569-2535. Disponível em: <<http://dx.doi.org/10.1016/j.infus.2015.09.005>>. Citado 2 vezes nas páginas 21 e 30.
- DEEDION, P. H. et al. Marketing meets web 2.0, social media, and creative commons: Implications for international marketing strategy. *Business Horizons*, v. 55, n. 3, p. 201-221, 2012. ISSN 0007-6243. SPECIAL ISSUE: STRATEGIC MARKETING IN A CHANGING WORLD. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0007624312000490>>. Citado na página 21.
- HONNAN, E. et al. Local e-government 2.0: Social media and corporate transparency in municipalities. *Government Information Quarterly*, v. 29, n. 2, p. 123-132, 2012. ISSN 0738-013X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0738013X1200010X>>. Citado na página 78.
- DUDIARDO, R. K. et al. Social crm features identification for higher education. *Journal of Engineering and Applied Sciences*, Medwell Journals, v. 12, n. 9, p. 2327-2331, 2017. Citado na página 81.

- CARR, C. T.; BAYES, R. A. Social Media: Defining, Developing, and Distorting. *Atlantic Journal of Communication*, v. 21, n. 1, p. 46-65, 2013. ISSN 15456880. Citado 2 vezes nas páginas 21 e 51.
- CARRASCAI, A. I. O. et al. Descobrimientos de conocimientos en historias clínicas mediante minería de textos. *RISIT-Serious Review de Sistemas e Tecnologías de Información*, Asociación Ibero de Sistemas e Tecnologías de Información (AISITI), n. 31, p. 29-41, 2019. Citado na página 53.
- CHAKRABORTY, A. et al. Who Makes Trends? Understanding Demographic Issues in Crowdsourced Recommendations. *in* *Proc.*, p. 22-31, 2017. Citado na página 51.
- CHANEY, A. J.; DUFF, D. M. Visualizing topic models. *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, p. 419-422, 2012. Citado 2 vezes nas páginas 28 e 29.
- CHARLTON-SMITH, L. E. et al. Using social media for actionable disease surveillance and outbreak management: A systematic literature review. *PLoS ONE*, v. 10, n. 10, p. 1-20, 2015. ISSN 1929-0333. Citado na página 71.
- CHEN, H.; HU, CHANG, H.; C. Soley, V. Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Quarterly*, v. 36, n. 4, p. 1165-1180, 2018. ISSN 01467366. Disponível em: <<http://www.jstor.org/stable/41700303>>. Citado na página 29.
- CHEN, Y. et al. Experimental explorations on short text topic mining between LDA and NMF based schemes. *Knowledge-Based Systems*, Elsevier B.V., v. 163, p. 1-13, 2019. ISSN 09507051. Citado 4 vezes nas páginas 28, 29, 36 e 54.
- CHINCHILLA, L. D. C. C.; FERREIRA, K. A. R. Analysis of the behavior of customers in the social networks using data mining techniques. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASNAAM 2016*, IEEE, p. 623-625, 2016. Citado 4 vezes nas páginas 27, 29, 51 e 54.
- CHIDHMALIA, K.; OCHAZI, P.; PAHIDA, V. Social media engagement strategy: Investigation of marketing and R&D interfaces in manufacturing industry. *Industrial Marketing Management*, v. 74, n. February 2017, p. 138-149, 2018. ISSN 00198501. Citado na página 53.
- CIRQUEIRA, D. et al. A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media. *in* 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 8-11 DEZ, 2018, p. 746-749. ISSN 978-1-5386-7225-6. Citado 4 vezes nas páginas 26, 52, 56 e 59.
- CIRQUEIRA, D. et al. Performance evaluation of sentiment analysis methods for Brazilian Portuguese. *Lecture Notes in Business Information Processing*, v. 263, p. 245-253, 2017. ISSN 18654390. Citado 2 vezes nas páginas 52 e 53.
- CIRQUEIRA, D. et al. Opinion Label: A Gamified Crowdsourcing System for Sentiment Analysis Annotation. *XVI Workshop de Ferramentas e Aplicações*, p. 205-213, 2017. Citado na página 43.
- CULOMO-PALACIOS, R. et al. Towards a social and context-aware mobile recommendation system for tourism. *Perucom and Mobile Computing*, Elsevier B.V., v. 38, p. 505-513, 2017. ISSN 15741152. Citado na página 29.

- CONSTANTINIDES, E.; HOLLESCHOVSKY, N. I. Impact of Online Product Reviews on Purchasing Decisions. *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, p. 271-278, 2016. Disponível em: <<http://www.scitegview.org/DigitalLibrary/Link.aspx?doi=10.5220/0007861002710278>>. Citado 7 vezes nas páginas 22, 25, 32, 34, 51, 67 e 83.
- CONSUMIDOR.GOV. Brazilian Consumer complaints: 2021. Disponível em: <<https://www.consumidor.gov.br>>. Citado 2 vezes nas páginas 82 e 84.
- COOPER, T.; STAVROS, C.; DOBIELE, A. R. Demises of influence: exploring negative sentiment in social media. *Journal of Product Brand Management*, 2019. ISSN 1063-0421. Disponível em: <<https://doi.org/10.1108/JPB-03-2018-1821>>. Citado na página 24.
- DAMANI, O. P. Improving Princeton Mutual Information (PMI) by incorporating significant co-occurrences. In: *CoNLL 2013 - 47th Conference on Computational Natural Language Learning, Proceedings*, SL: s.n., 2013, p. 20-28. ISSN 9781947264793. Citado na página 70.
- DOMO. Data Kiter Steps 7.8. 2019. 1 p. Disponível em: <<https://www.domo.com/tech/data-never-steps-7>>. Citado na página 26.
- DUNN, K.; HARNISS, D. Whose voice is heard? The influence of user-generated versus company-generated content on consumer expectations towards CSR. *Journal of Marketing Management*, Routledge, v. 35, n. 9-10, p. 880-915, 2019. ISSN 14721176. Disponível em: <<https://doi.org/10.1080/0267257X.2019.1666901>>. Citado na página 22.
- DYSON, H. et al. Evaluating the use of Facebook to increase student engagement and understanding in lecture-based classes. *Higher Education*, Elsevier Academic Publishers, v. 60, n. 3, p. 303-312, feb 2015. ISSN 00181560. Disponível em: <<https://link.springer.com/article/10.1007/s10734-014-9776-3>>. Citado na página 78.
- EASLEY, D.; KLEINBERG, J. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. [s.n.], 2010. 21-46 p. Disponível em: <<http://www.mcs.well.ox.ac.uk/people/kleinber/networks-book/>>. Citado na página 83.
- EBEL. 43º Webshoppers. p. 1-80, 2021. Citado na página 25.
- EINWILLER, S. A.; STELLEN, S. Handling complaints on social network sites - An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies. *Public Relations Review*, Elsevier Inc., v. 41, n. 2, p. 186-201, 2015. ISSN 0363-8111. Citado 2 vezes nas páginas 51 e 58.
- ELNAGA, S. K. et al. Characterizing audience engagement and assessing its impact on social media disclosure of mental illness. In: *Fifth International AAAI conference on web and social media*, [S.L.: s.n.], 2018. Citado na página 73.
- FAARI, H.; HELMS, R.; SPIRIT, M. Web 2.0 in the CRM domain: defining social CRM. *International Journal of Electronic Customer Relationship Management*, v. 5, n. 1, p. 1, 2011. ISSN 1750-0664. Disponível em: <<http://www.intellectbooks.com/doi.php/51-2011>>. Citado 2 vezes nas páginas 22 e 23.

- FANG, A. et al. Examining the coherence of the top ranked tweet topics. *SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 825-828, 2016. Citado 3 vezes nas páginas 68, 69 e 71.
- FANG, A. et al. Examining the coherence of the top ranked tweet topics. *SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 825-828, 2016. Citado na página 71.
- FANG, B. et al. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management, Elsevier Ltd.*, v. 52, p. 498-506, 2016. ISSN 02615177. Disponível em: <<http://dx.doi.org/10.1016/j.tourman.2015.07.018>>. Citado na página 65.
- PARSIEV, A.; CHUA, T. S. TweetFit: Fusing multiple social media and sensor data for wellness profile learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, p. 95-101, 2017. Citado na página 25.
- FERNANDES, L. C. et al. An extensive analysis of online restaurant reviews: A case study of the Amazonian Culinary Tourism. *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, FedCSIS 2020*, v. 21, p. 91-94, 2020. Citado 2 vezes nas páginas 83 e 86.
- FLESCH, R. F. A new readability yardstick. *The Journal of applied psychology*, v. 32 3, p. 221-30, 1948. Citado 2 vezes nas páginas 69 e 81.
- GLAUCIA, S. et al. *Data Preprocessing in Data Mining [S.L.: etc.]*, 2015, 105-243 p. ISSN 16094008. ISBN 9783319100107. Citado na página 83.
- GAVILANES, J. M.; FLATTEN, T. C.; BRUTHER, M. Content Strategies for Digital Consumer Management in Social Networks: Why Advertising Is an Antecedent of Engagement. *Journal of Advertising*, v. 47, n. 1, p. 4-23, 2018. ISSN 00474037. Citado 12 vezes nas páginas 75, 22, 26, 31, 32, 40, 41, 42, 46, 47, 51 e 58.
- GENCAYEH, N.; ABBAN, A. A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software, Elsevier Inc.*, v. 125, p. 207-219, 2017. ISSN 01651212. Citado na página 71.
- GLOBALWIRENUEX. *Brazilian Consumers: Social Media Habits and Consumption Views - GWI [S.L.]*, 2020. Disponível em: <<https://www.globalwiredna.com/regions/brazilian-consumers>>. Citado na página 25.
- GOSHIZ, M. et al. A Recommender System of Buggy App Clusters for App Store Moderations. *Proceedings - 2nd ACM International Conference on Mobile Software Engineering and Systems, MOBILESoft 2015*, p. 1-11, 2015. Citado na página 71.
- GREENBERG, P. *Social crime scenes of age*. Reviewed by Oracle, 2008. Citado 2 vezes nas páginas 28 e 80.
- GREENBERG, P. *CRM at the Speed of Light: Social CRM Strategies, Tools, and Techniques*. [S.L.: McGraw-Hill New York, 2010. Citado na página 28.

- GRETZEL, U. et al. Conceptual foundations for understanding smart tourism ecosystems. *Computers in Human Behavior*, Elsevier Ltd, v. 50, p. 328-343, 2015. ISSN 0747-5620. Citado na página 39.
- HAN, H. J. S. et al. What guests really think of your hotel: Text analysis of online customer reviews. *Cornell Hospitality Report*, v. 16, n. 2, p. 3-17, 2016. Disponível em: <<http://ehospitality.cornell.edu/chreporter>>. Citado na página 53.
- HARRIGAN, P. et al. Customer engagement with tourism social media brands. *Tourism Management*, v. 59, p. 507-509, 2017. ISSN 0263-5777. Citado na página 25.
- HE, L. et al. The voice of drug consumers: Online textual review analysis using structural topic model. *International Journal of Environmental Research and Public Health*, v. 17, n. 10, 2020. ISSN 1660-4601. Citado na página 82.
- HIGUCHI, M. et al. Grouping ecometrics: a systematic review of information available on the Internet. *American Journal of Obstetrics and Gynecology*, Elsevier, v. 216, n. 5, p. 451-458 e1, 2017. ISSN 1097-6868. Citado 3 vezes nas páginas 66, 70 e 82.
- HILASTINSKI, S.; ACHABE, N. M. How are campus students using social media to support their studies? an exploratory interview study. *Education and Information Technologies*, Springer, v. 17, n. 4, p. 451-464, 2012. Citado na página 76.
- HENDIC, A. The transformation of higher education: evaluation of crm concept application and its impact on students satisfaction. *European Business Review*, Springer, v. 6, n. 1, p. 53-77, 2016. Citado na página 20.
- HUSSAIN, S. et al. Consumers' online information adoption behavior: Motives and antecedents of electronic word of mouth communications. *Computers in Human Behavior*, Elsevier Ltd, v. 80, p. 22-32, 2018. ISSN 0747-5620. Citado na página 38.
- IBGE. Estimativas da população com referência a 1º de julho de 2019 (16). 2019. Disponível em: <<https://agenciademecias.ibge.gov.br/agencia-detalhes-da-unidade/?view=medialibrary&id=216&id=2000>>. Citado 4 vezes nas páginas 13, 24, 57 e 62.
- IBGE. PNAD continua - Pesquisa Nacional por Amostra de Domicílios Contínua (53). 2019. 8 p. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101290_informativo.pdf>. Citado na página 24.
- IBGE - Instituto Brasileiro de Geografia e Estatística. Produto Interno Bruto - PIB | IBGE 2020. Disponível em: <<https://www.ibge.gov.br/explica/ PIB.php>>. Citado 3 vezes nas páginas 15, 58 e 82.
- INEP. Brazilian Higher Education Census 2021. Disponível em: <<https://www.gov.br/inep/pl-br/area-de-educacao/percepcao-estatisticas-e-indicadores/censo-da-educacao-superior/resultados>>. Citado 2 vezes nas páginas 32 e 34.
- JARVIS, J. My Doll built. *The Guardian*, 2011. Disponível em: <<https://www.theguardian.com/technology/2015/jun/29/unnecessaryinnovation.blogging>>. Citado na página 78.
- JO, J. M.; PEREIRA, M. An evaluation of sentiment analysis for mobile devices. 11 October, 2017. Citado na página 82.

- KAPLAN, A. M.; HAENSLIN, M. Users of the world, unite! The challenges and opportunities of Social Media. 2010. Citado na página 21.
- KARNA, N.; SUPRIANA, I.; MAULIDEVI, N. Social crm using web mining for Indonesian academic institution. In: *IEEE 2015 International Conference on Information Technology, Systems and Innovation (ICITSI)*. [S.l.], 2015. p. 1-6. Citado 2 vezes nas páginas 78 e 81.
- KEMP, S. *Digital 2021 - We Are Social 2021*. 8 p. Disponível em: <<https://wearesocial.com/digital-2021/>>. Citado na página 25.
- KIM, A. J.; JOHANSON, K. E. Power of consumers using social media: Examining the influence of brand-related user-generated content on Facebook. *Computers in Human Behavior*, Elsevier Ltd, v. 58, p. 96-108, 2016. ISSN 07475632. Disponível em: <<https://dx.doi.org/10.1016/j.chb.2015.12.047>>. Citado 3 vezes nas páginas 24, 32, 53 e 67.
- KUHNIA, M.; LENDEL, V. Successful Application of Social CRM in The Company. *Procedia Economics and Finance*, Elsevier, v. 23, p. 1190-1194, Jan 2015. ISSN 2212-5671. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2212567115000676>>. Citado na página 22.
- KUMAR, A.; DABAS, V.; BHODA, P. Text classification algorithms for mining multistructured data: a SWOT analysis. *International Journal of Information Technology*. Springer Singapore, 2018. ISSN 2511-2294. Disponível em: <<https://doi.org/10.1007/s41070-017-0072-1>>. Citado na página 35.
- LI, D. et al. Integration of Knowledge Graph Embedding Into Topic Modeling with Hierarchical. In: *Co*, p. 940-950, 2019. Citado 3 vezes nas páginas 28, 39, 53 e 71.
- LI, H. Traditional to hybrid: Social media's role in reshaping innovation in higher education. In: *Digital Arts and Entertainment: Concepts, Methodologies, Tools, and Applications*. [S.l.]: IGI Global, 2014. p. 287-311. Citado na página 78.
- LI, Y. et al. The concept of smart tourism in the context of tourism information services. *Tourism Management*, Elsevier Ltd, v. 58, p. 291-300, 2017. ISSN 0969-1177. Citado na página 39.
- LIU, B. et al. Growing story forest online from massive breaking news. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. [S.l.], 2017. p. 777-785. Citado na página 61.
- LIU, J. et al. Like it or not: The Fortune 500's Facebook strategies to generate users' electronic word-of-mouth. *Computers in Human Behavior*, Elsevier Ltd, v. 74, p. 600-613, 2017. ISSN 07475632. Citado na página 54.
- LORETO, P. et al. Social CRM: Biggest challenges to make it work in the real world. *Lecture Notes in Business Information Processing*, v. 263, p. 211-221, 2017. ISSN 1865-1348. Citado 7 vezes nas páginas 21, 23, 39, 51, 67, 82 e 83.
- MAIZ, A.; ARRANZ, N.; JUAN, J. C. Factors affecting social integration on social network sites: the Facebook case. *Journal of Enterprise Information Management*, v. 29, n. 5, p. 630-649, 2016. ISSN 17510398. Citado 2 vezes nas páginas 37 e 41.

- MAROLT, M.; ZIMMERMANN, H. D.; PUCHHAIR, A. Exploratory study of social CRM use in SMEs. *Engineering Economics*, v. 29, n. 4, p. 468-477, 2019. ISSN 14022753. Citado na página 23.
- MARSHALL, A.; MUECK, S.; SHOCKLEY, R. How leading organisations use big data and analytics to innovate. *Strategy and Leadership*, v. 43, n. 5, p. 37-48, 2015. ISSN 10678572. Citado na página 23.
- MCHERTY, S. et al. Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Software Engineering, Empirical Software Engineering*, v. 21, n. 3, p. 1067-1106, 2016. ISSN 15737616. Citado 2 vezes nas páginas 32 e 33.
- MEYLIANA, P.; HIDAYANTO, A. N.; BULHARDO, E. K. Social media adoption for social crm in higher education: An insight from Indonesian universities. *International Journal of Synergy and Research*, Wydawnictwo Uniwersytetu Maria Curie-Skłodowskiej, v. 4, n. 2, 2015. Citado na página 40.
- MIRAF, M. da Justiça e seguradora pública. *Data from: Consumidor.gov.br - Dados do Brazilian Open Data Portal*, 2021. Disponível em: <<https://dados.gov.br/dataset/reclamacoes-do-consumidor-gov-br>>. Citado na página 32.
- MUJAHID, S. et al. An empirical study of Android Wear user complaints. *Empirical Software Engineering, Empirical Software Engineering*, v. 23, n. 6, p. 3176-3202, 2018. ISSN 15737616. Citado na página 50.
- NAIR, C.; CHAN, S.; FANG, X. A case study of crm adoption in higher education. In: CHEN, H. *Proceedings of the 2007 Information Resources Management Association International Conference*. [S.l.], 2007. Citado na página 79.
- NETSCA. TOP ECOMMERCE RANKING REPORTS. Disponível em: <https://ecommerce-brasil.rankings.netquest.digital>. Acesso em: 27/01/2020, 2020. Disponível em: <<https://ecommerce-brasil.rankings.netquest.digital/>>. Citado na página 60.
- NETTO, E. N. D. et al. Melhorando sistemas de social crm por meio de eletrônica word-of-mouth. *Revista Eletrônica de Iniciação Científica em Computação*, v. 17, n. 4, 2019. Citado na página 23.
- NGUYEN, D. T. et al. Rapid Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks. n. *Isosm*, p. 652-655, 2016. Disponível em: <<http://aisel.net/dm/1008.0302>>. Citado 2 vezes nas páginas 28 e 29.
- NOGUEIRA DE SOUSA, G. et al. Adoption of Social CRM in Micro and Small Enterprises in Analysis of Narahin's Market. *Proceedings of the 15th CONTECSI International Conference on Information Systems and Technology Management*, v. 15, p. 0-2, 2018. Citado 2 vezes nas páginas 30 e 41.
- NUSARI, K. et al. A theoretical framework of electronic word-of-mouth against the backdrop of world networking activities. *Journal of Travel and Tourism Marketing*. Routledge, v. 34, n. 5, p. 653-665, 2017. ISSN 10546606. Citado na página 34.

- OLIVEIRA, B.; CASARIS, B. The importance of user-generated photos in restaurant selection. *Journal of Hospitality and Tourism Technology*, 2018. ISSN 17579891. Cadeia 2, versão não paginas 24 e 26.
- OLIVEIRA, L. Social students relationship management in higher education: extending educational and organizational communication into social media. In: 9th Annual International Technology, Education and Development Conference, (ATED), [S.l.: s.n.], 2015. Citado na página 80.
- OLMEDILLA, M.; MARTÍNEZ-TORRES, M. R.; TORAL, S. Harvesting big data in social science: A methodological approach for collecting online user-generated content. *Computer Standards & Interfaces*, Elsevier, v. 36, p. 79-87, 2014. Citado na página 32.
- ORENGA-BOULA, S.; CHALMEIDA, R. Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies. SpringerPlus, Springer International Publishing, v. 5, n. 1, 2016. ISSN 21931801. Cadeia 3 versão não paginas 23, 24 e 67.
- OTHMAN, I. W. et al. Text readability and fraud detection. (ISDEIA 2012 - IEEE Symposium on Business, Engineering and Industrial Applications, IEEE, n. 99, p. 296-301, 2012. Citado 2 vezes não paginas 70 e 81.
- PARK, B. O.; CHAE, H. K.; KWON, J. The structural topic model for online review analysis: Comparison between green and non-green restaurants. *Journal of Hospitality and Tourism Technology*, v. 11, n. 1, p. 1-17, 2018. ISSN 17579891. Citado na página 83.
- PEDREGONZA, F. et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, JMLR.org, v. 12, p. 2825-2830, 2011. Citado na página 43.
- PODITO, J. S. et al. Legibilidade de artigos de um periódico acadêmico na área de melhoramento vegetal. *Culturas e Saber*, v. 7, n. 2, p. 205-211, 2011. Citado na página 74.
- PRADIPYADINI, C. Social Media Marketing: Measuring Its Effectiveness and Identifying the Target Market. *Journal of Undergraduate Research*, v. 14, p. 1-11, 2011. Citado na página 30.
- PIRESS, G. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, *Survey Says: Forbes Tech*, p. 4-5, 2016. Disponível em: <<https://www.forbes.com/sites/glipress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=061ce5280863> > <https://www.forbes.com/sites/glipress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>>. Citado na página 28.
- QI, J. et al. An overview of data fusion techniques for Internet of Things enabled physical activity recognition and posture. *Information Fusion*, Elsevier B.V., v. 55, n. September 2019, p. 209-260, 2020. ISSN 15692501. Citado na página 82.
- QUATTRONE, G. et al. Is the sharing economy about sharing at all? A linguistic analysis of Airbnb reviews. In: 11th International AAAI Conference on Web and Social Media, (IWSM 2018) [S.l.: s.n.], 2018, p. 660-671. ISBN 9781577357968. Citado na página 48.
- RAMAN, R.; MENON, P. Using social media for international market segmentation of family firms. *International Journal of Innovation Science*, 2018. ISSN 15472231. Citado na página 34.

SALMINEN, J. et al. Anatomy of Online Hate: Developing a Taxonomy and Machine Learning Models for Identifying and Classifying Hate in Online News Media. *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*, n. brown, p. 330-339, 2018. Citado na página 56.

SASRI, C.; BRYNILDSEN, G.; BILCHAN, A. Social media, customer engagement and advocacy. *International Journal of Contemporary Hospitality Management*, p. LICHM-02-2018-0108, 2019. ISSN 0950-6119. Disponível em: <<https://www.emeraldinsight.com/doi/10.1108/LICHM-02-2018-0108>>. Citado na página 24.

SCHAFERIC, F. et al. Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes. *2019 IEEE International Conference on Technology Management, Operations and Decisions, ICTMOD 2019*, p. 190-195, 2019. Citado 4 vezes nas páginas 27, 28, 52 e 53.

SCHMAH, M.; WILKE, T.; ROSSMANN, A. Electronic Word-of-Mouth: A Systematic Literature Analysis. *Lecture Notes in Informatics (LNI)*, p. 147, 2017. Citado 4 vezes nas páginas 22, 39, 51 e 67.

SKEMAN, E. D.; O'HARA, M. Customer relationship management in higher education: Using information systems to improve the student-staff relationship. *Campus-wide information systems*. Emerald Group Publishing Limited, 2000. Citado na página 30.

SHARMA, R. et al. Digital literacy and knowledge societies: A grounded theory investigation of sustainable development. *Telecommunications Policy*, v. 40, n. 7, p. 628-643, 2014. ISSN 03085961. Citado na página 52.

SHIAU, W. L.; DWIVEDI, Y. K.; LAI, H. H. Examining the user knowledge on facebook. *International Journal of Information Management*, Elsevier, v. 43, n. December 2017, p. 62-63, 2018. ISSN 0288-012. Disponível em: <<https://doi.org/10.1016/j.ijinfman.2018.06.000>>. Citado 2 vezes nas páginas 21 e 29.

SILVA, E. B. D. et al. Redes sociais e permissões da mídia: utilização do facebook no contexto da direção do sungus. *RISTI: Revista Brasileira de Sistemas e Tecnologias de Informação*. Associação Brasileira de Sistemas e Tecnologias de Informação (ABSTI), n. 30, p. 107-122, 2018. Citado na página 51.

SILVA, W. et al. A methodology for community detection in Twitter. *Proceedings of the International Conference on Web Intelligence - WI '17*, p. 1086-1089, 2017. Citado 2 vezes nas páginas 30 e 67.

Simão Kemp. *Digital 2021: Brasil/2021*. Disponível em: <<https://datareportal.com/reports/digital-2021-brasil>>. Citado na página 25.

SILVA, G. N. de; JUNIOR, A. F. L. J.; LEMAYO, F. M. F. Facebook posts for analyzing Content Strategies for Digital Consumer Engagement: a curated dataset. *Zenodo*, 2021. Disponível em: <<https://doi.org/10.5201/zenodo.541286>>. Citado 2 vezes nas páginas 42 e 66.

STATISTA. • *Global social media ranking 2019* / Statista, 2019. Citado na página 34.

STONE, M. The evolution of the telecommunications industry-What can we learn from it? *Journal of Direct, Data and Digital Marketing Practice*, v. 16, n. 3, p. 157-165, 2015. ISSN 17460174. Cited on page 37.

TANG, C.; GUO, L. Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication. *Marketing Letters*, Springer, v. 26, n. 1, p. 67-80, 2015. Cited on page 33.

THIRUNILLAI, S.; TELLES, G. J. Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, v. 31, n. 2, p. 188-215, 2012. ISSN 07322399. Cited on page 25.

TRSTENJAK, U.; MIKAC, S.; DONKO, D. RNN with TF-IDF based framework for text categorization. *Procedia Engineering*, Elsevier B.V., v. 69, p. 1156-1164, 2014. ISSN 18770665. Cited on page 56.

VEECHIO, P. D. et al. Creating value from Social Big Data: Implications for Smart Tourism Destinations. *Information Processing and Management*, Elsevier, v. 54, n. 5, p. 847-860, 2018. ISSN 03064573. Cited on page 24.

VERMEER, S. A. et al. Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, Elsevier B.V., n. xxxx, p. 1-17, 2019. ISSN 0020-118X. Cited 4 times on pages 32, 31, 67 e 82.

VII, P. M. et al. Mining user opinions in mobile app reviews: A keyword-based approach. *Proceedings - 2015 IEEE/ACM International Conference on Automated Software Engineering, ASE 2015*, IEEE, p. 719-726, 2016. Cited 2 times on pages 32 e 31.

VYDINWARAN, V. G. et al. "Bacon bacon bacon": Food-related tweets and sentiment in metro deodor. In: *2018 International AAAI Conference on Web and Social Media, ICWSM 2018*, [N.L. s.R.], 2018. p. 402-405. ISSN 9781577357968. Cited 2 times on page 68 e 69.

WALL, A. P.; WRIGHT, L. T. Customer relationship management and service quality: Influences in higher education. *Journal of Customer Behaviour*, Western Publishers Ltd, v. 15, n. 1, p. 67-79, 2016. Cited on page 41.

WANG, H. et al. Review on mining data from multiple data sources. *Pattern Recognition Letters*, v. 109, p. 120-128, 2018. ISSN 01678655. Cited on page 25.

WANG, Y.; YU, C. Social interaction-based customer decision-making model in social commerce: The role of word of mouth and observational learning. *International Journal of Information Management*, v. 37, p. 179-189, 2015. ISSN 02664012. Cited on page 22.

WHITING, A.; WILLIAMS, D. Why people use social media: a new and gratifications approach. *Qualitative Market Research: An International Journal*, v. 16, n. 4, p. 362-369, 2013. ISSN 13522752. Cited on page 21.

WHITE, R. CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, n. 2000, p. 29-39, 2000. Cited 8 times on pages 12, 27, 28, 29, 33, 54, 82 e 83.

- WITTWER, M.; REINHOLD, O.; ALT, R. *Customer Context and Social CRM: A Literature Review and Research Agenda*. 30th Biol eConference Digital Transformation From Connecting Things to Transforming Our Lives, p. 1-14, 2017. Citado na página 21.
- XIANG, Z. et al. A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. *Tourism Management*, Elsevier Ltd, v. 58, p. 31-65, 2017. ISSN 0261-5177. Citado 4 vezes nas páginas 67, 68, 69 e 71.
- ZHANG, J. Multi-source remote sensing data fusion: Status and trends. *International Journal of Image and Data Fusion*, v. 1, n. 1, p. 5-20, 2010. ISSN 1947-8621. Citado na página 83.
- ZHANG, L.; YANG, W. Consumers' responses to invitations to write online reviews. *International Journal of Contemporary Hospitality Management*, p. IJCHM 01-2018-0022, 2018. ISSN 0959-6119. Disponível em: <<https://www.emeraldinsight.com/doi/10.1108/IJCHM-01-2018-0022>>. Citado na página 24.

APÉNDICES

As aplicações públicas e estudos de dados estão associados de forma íntima, algoritmos e bases de dados específicas para atingir os seus objetivos. Isso é observado ao comparar os trabalhos de [17, 18, 19], que usam o Teorema do Teorema como base de dados principal, no entanto com algoritmos e propriedades diferentes. Além disso, outros estudos utilizam as bases de dados de factcheck, tal como [20], que compara diferentes algoritmos de extração de informações em conteúdos. Ainda no contexto de análise de sentimentos, em [21] experimentamos uma plataforma de *crowdsourcing* para a obtenção de informações. Por fim, uma metodologia de *big data* para análise de sentimentos baseada em uma plataforma online destinada ao reconhecimento de sentimentos de empresas é utilizada em [12].

3 METODOLOGIA

Este trabalho tem um componente derivativo da literatura sobre análise de dados em si, com o objetivo de analisar o estado da arte e da prática sobre análise de dados sociais, a partir da identificação de ferramentas, bases de dados, algoritmos e de estratégias de uso de dados públicas citadas [22, 23].

Para a construção desta metodologia derivativa foram usadas as recomendações propostas por [24] e [25]. Com base nisso, este estudo foi conduzido em três etapas: Planejamento, Construção, e Avaliação. As definições e atividades referentes ao planejamento e construção podem ser observadas nas subseções abaixo. O Relatório é detalhado no Seção 4.

3.1 Planejamento

No início do planejamento ocorre a identificação de etapas de análise que por meio da definição das perguntas de pesquisa, definem as hipóteses as bases de dados e fontes consideradas e critérios de seleção dos dados [26]. A identificação destas atividades para o presente trabalho encontra-se descrita a seguir:

3.1.1 Perguntas de pesquisa. As perguntas de pesquisa definem o objetivo de cada etapa quando se trata de análise de dados sociais. Para cada método qualitativo ou quantitativo usado, deve-se considerar suas limitações e métodos utilizados acerca dos tipos de dados coletados e como eles se relacionam com publicações de análise de dados sociais. As perguntas de pesquisa deste trabalho são apresentadas a seguir:

- PP1: Quais são as principais bases de dados para análise de dados sociais?
- PP2: Quais são as principais técnicas e algoritmos para análise de dados sociais?
- PP3: Como as técnicas apresentadas pela literatura sobre análise de dados sociais estão relacionadas?

Para responder essas questões, os dados serão de literatura, a saber: bases de dados sobre métodos (PP1), técnicas e algoritmos mais prevalentes (PP2) e quais as aplicações e bases relevantes (PP3).

3.1.2 Processo de Pesquisa. Há uma dificuldade de encontrar fontes relevantes e consideráveis a uma aplicabilidade de análise de dados sociais e de acessibilidade de se encontrar as fontes relevantes, então se faz necessário que trabalhos sobre análise de dados sociais sejam analisados para a base de análise de dados sociais. Os critérios para fontes selecionadas observados e observados

Tabela 1: Critérios de inclusão e exclusão de trabalhos.

Critérios de inclusão:	Trabalhos completos
	Trabalhos publicados em idiomas que se possa ler
	Trabalhos sobre análise de dados sociais
Critérios de exclusão:	Trabalhos não científicos ou não acadêmicos
	Trabalhos sobre análise de dados sociais em outras áreas
	Trabalhos não publicados antes de 1970 e após 2019

selecionados através de critérios. Sobre as fontes consultadas, Joseph Schuler (2018)¹, afirma, tendo em consideração as seguintes características, que estas referências são fontes de dados em análise social: KWOM, WISOM, COWOM e WOM (18).

3.1.3 Seleção dos dados. A fim de selecionar somente trabalhos relevantes ao tema de análise de dados sociais, a seleção dos dados baseia-se nos critérios de inclusão e exclusão apresentados na Tabela 1. Como critérios de inclusão foram considerados todos os trabalhos completos publicados em idiomas que se possa ler sobre análise de dados sociais. Por outro lado, foram excluídos artigos e artigos de dados, revisões sistemáticas, trabalhos que não fossem análise de dados sociais e trabalhos que não utilizassem métodos base de dados para análise de dados sociais. Por se considerar isso, foi que se analisou as citações de 2019 da publicação WOM IT, foram publicados durante a construção deste trabalho, portanto, foi excluído base de dados após a publicação.

3.2 Construção

A etapa de construção ocorreu na criação dos dados definidos na base de planejamento. A seguir, alguns aspectos técnicos são descritos:

3.2.1 Criação dos trabalhos. As fontes foram consultadas na primeira análise pública de trabalhos em diferentes plataformas. Para isso, foi necessário a identificação de diferentes fontes de dados e métodos para a coleta de alguns dados disponíveis em plataformas, como: Twitter, Web, Youtube e, por alguns casos, pesquisas online. Entre dados coletados foram considerados os que estavam sobre fontes literárias aplicadas ou os critérios de inclusão e exclusão e, por fim, ocorreu a publicação que está dada nas referências selecionadas foram utilizadas para planejamento adicional.

3.2.2 Análise dos trabalhos. Os trabalhos literários foram analisados através de uma análise derivativa, a fim de identificar ferramentas, bases de dados e algoritmos utilizados. Para isso, os trabalhos foram analisados considerando suas palavras-chave. Para isso, foi considerado uma base de palavras-chave com o objetivo de analisar a construção das técnicas e recomendações sobre a análise dos trabalhos analisados neste estudo.

4 RESULTADOS E DISCUSSÃO

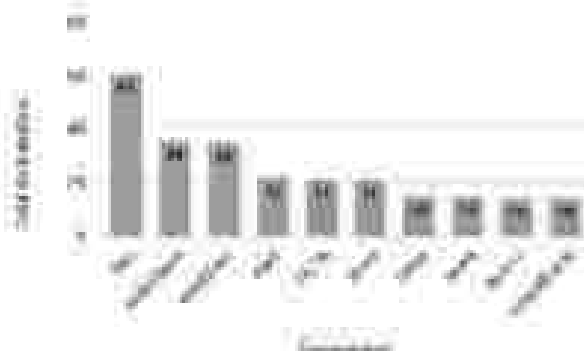
O processo de coleta dos trabalhos resultou em 164 trabalhos publicados nos últimos meses após a criação das fontes literárias. Desta, as citações de trabalhos e métodos foram selecionadas para a construção

¹<http://www.schuler.com/2018/04/04/10-social-media-data-sources/>

Tabela 3. Ferramentas mais utilizadas e mais desenvolvidas

Ferramenta	Descrição
LWS	Um conjunto diversificado de ferramentas de 1999, <i>Supply and Word Tools (LWS)</i> é um instrumento de análise de texto com mais de 4.000 palavras de língua inglesa categorizadas em seus domínios. Desenvolvido para o objetivo de apoiar estudos sociais e psicológicos de nível de texto, a ferramenta é muito utilizada para estudos de processamento de linguagem natural [41].
nltools	Ferramenta de aprendizagem de máquina desenvolvida em Python para visualização de dados e análise de texto. Seu uso está associado a tarefas como: processamento, redução de dimensionalidade, classificação, regressão e clustering em dados também categorizados, como palavras [42].
Word2Vec	Um conjunto de técnicas associadas a tarefa de word embedding, responsáveis por associar cada palavra presente em um corpus a uma representação vetorial e, desta maneira, para aplicações de processamento de linguagem natural [43].
MITI	Amazon Mechanical Turk (MTurk) é uma plataforma de crowdsourcing que permite a um de seus fornecedores distribuída para a execução de tarefas de processamento, como redução de dados e respostas a uma pesquisa. A comparação interna dos indivíduos envolvidos no processo permite a criação de trabalhos de redução e análise de dados [44].
NLTK	Natural Language Toolkit (NLTK) é uma plataforma para processamento e análise de linguagem natural, oferecendo suporte a nível de texto, desde técnicas como redução, clustering, PPT, regressão, corpus, léxico, entre outros [45].
CLIM	Código para o Projeto de Processamento de Linguagem Natural da Universidade de Stanford, <i>Stanford Metrics for Word Representations (CLIM)</i> é um conjunto de representações vetoriais, mas especialmente para aplicações de representação vetorial de palavras [46, 47].
Core	Projeto de análise de texto disponibilizado no GitHub [48]. Trata-se uma plataforma para Deep Learning escrita em Python [49]. Uma série API de fácil compreensão utilizando por meio de <i>YML</i> as variáveis e respostas como <i>Naïve</i> e <i>Text</i> . Core é a segunda ferramenta mais utilizada para Deep Learning no mundo [50].
MLDA	Machine Environment for Knowledge Analysis (MLDA) é uma ferramenta computacional para propagação, classificação, regressão, clustering, associação e redução de dados [51].
Text4	Ferramenta que permite a busca, comparação e delimitação de texto e partes de corpo humano [52].
CountServer	Plataforma de crowdsourcing com mais de 1 milhão de trabalhadores voluntários [53]. Nela, os trabalhadores podem ser recrutados por meio de comunicação direta, por meio de uma principal [54], ou os trabalhadores podem aceitar as tarefas com base na recompensa, redução de tarefas, redução de redução dos trabalhadores [55].

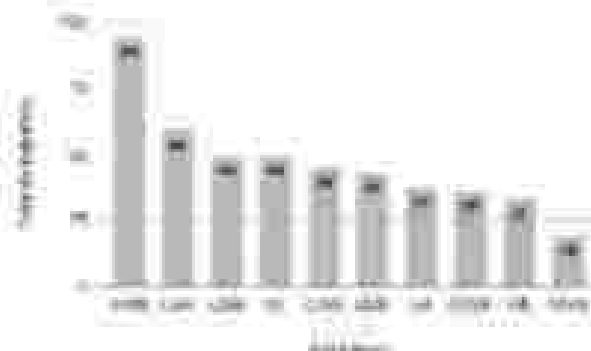
Figura 3. As dez ferramentas mais utilizadas nos trabalhos acadêmicos



MITI = Amazon Mechanical Turk; NLTK = Natural Language Toolkit

Resposta: as redes neurais, pode-se responder à PPT: "Quais são as principais técnicas e algoritmos nos estudos de análise de texto?" Os resultados obtidos mostram a existência e predominância de ferramentas que atuam no campo do processamento de texto, como observado nos 10 principais trabalhos de palavras listados na Figura 3. Sobre estas ferramentas, algumas conclusões foram de:

Figura 4. Os dez algoritmos mais utilizados nos trabalhos acadêmicos



MITI = Amazon Mechanical Turk; NLTK = Natural Language Toolkit; Word2Vec = Stanford Metrics for Word Representations; Core = Stanford Metrics for Word Representations; CLIM = Stanford Metrics for Word Representations; CountServer = Stanford Metrics for Word Representations; LWS = Stanford Metrics for Word Representations; NLDA = Stanford Metrics for Word Representations; Text4 = Stanford Metrics for Word Representations; Word2Vec = Stanford Metrics for Word Representations

em um ponto de vista. A ferramenta mais utilizada é a *Supply and Word Tools (LWS)*, em um ponto de vista, pode-se justificar pelo fato de que possui um grande número de palavras em inglês

- [illegible]

APÊNDICE B – ARTIGO PUBLICADO NO JCRM (2021) – SOCIAL CRM AS A BUSINESS STRATEGY: DEVELOPING DYNAMIC CAPABILITIES OF MICRO AND SMALL ENTERPRISES

Social CRM as a business strategy: developing dynamic capabilities of Micro and Small Enterprises

Isabella de Silva Guimarães¹, Gustavo Viegas de Sousa², Anderson José Junior³ and
 Fábio Manuel França Lobo^{1*}

¹ Federal University of Viçosa - Brazil
² Nova University of Madrid - Spain
 Email: isilva@ufv.br

Abstract. The global pandemic caused by the spread of COVID-19, has transformed the way people go through shopping in the same. Social Media is an important channel for sharing information about products, services, and brands. Thus, considering the number (distance) of these channels is this new channel, this paper describes the results of a survey to assess the Social Media to improve customer relationship management processes in 11 companies. The focus was on digital marketing for micro and small enterprises. Moreover, we discussed first impressions in-depth, identifying challenges and perspectives on Social CRM adoption by micro and small enterprises. The results show that this is still an exceptional environment for micro and small companies, but with great potential to boost sales, develop customer loyalty and increase brand visibility. The lessons learned have the potential of guiding policymakers to create more adequate measures for boosting this market sector.

Keywords: Digital Marketing; Micro and Small Enterprises; Social CRM.

1. Introduction

The global pandemic caused by the spread of COVID-19, has transformed the way people shop [1]. Companies adapted their purchases to e-commerce platforms, a fact that positions the mobility as one of the main pillars on sustaining the global economy [2]. According to [3], the post-consumers' trade will no longer be the same. Therefore, it is necessary for companies to set time interventions to understand their client in the best possible way [4]. In this scenario, Social Media (SM) platforms stand out. In January 2021, SM platforms reported over 4.2 billion active users [5]. It is noteworthy that this was the biggest growth in three years. This growing phenomenon of digital generation has become one of the most fundamental ones when it comes to product development and trends by retail, fashion, and large companies [6]. From User-Generated Content on social media, companies are able to make predictions about new trends and remain competitive in their market niche [7]. However, it is not feasible to remain in traditional customer management strategies [8]. Hence, it was necessary to alter traditional Customer Relationship Management (CRM) processes for integrated in marketing, innovation, sales, user experience, and other sales, to the CRM environment [9].

Social CRM is the adoption of interaction between companies and customers through social media. It can also be defined as a set of processes that make use of technologies to capture and analyze customer data in order to know their individual needs [10]. Moreover, with the knowledge generated from these data companies are able to understand the behavior of their customers and objectively identify their needs [11]. In addition, marketing campaigns can be more efficient in order to directly reach their target audience and convert leads into buyers [12]. This scenario is valid for both large companies and Micro and Small Companies (MSCs) [13]. It is important to highlight the representativeness of MSCs in the Brazilian economy. Currently, small businesses account for 99% of the 8.4 million Brazilian companies and are responsible for about 12% of formal jobs in the private sector [14].

Despite this economic relevance, this sector reports a shortage of qualified professionals to fulfill their demands and they are expensive to implement Social CRM strategies in their businesses. It was identified in the existing literature that few studies are directed to study the use of Social CRM strategies by MSCs. Therefore, based on its economical relevance for the country, it is necessary to analyze the implementation of Social CRM strategies by Micro and Small Companies in order to identify opportunities for intervention. In these gaps, the following research questions defined: RQ1: Which aspects of Social CRM are important for MSCs? RQ2: How do MSCs adopt Social CRM strategies in their businesses?

To answer these questions we conducted a 3-step investigation on how to use Social Media to improve customer relationship management, identifying challenges and perspectives on Social CRM adoption by micro and small enterprises. The first step consisted of applying a survey on 31 companies, aiming to identify some quantitative aspects. In the second step, we collected and analyzed five case studies companies from different areas in the market. We also highlighted some lessons learned and perspectives that have the potential of guiding practitioners to create more adequate measures for boosting the MSCs sector. The remainder of this paper is structured as follows. Section 2 presents a brief survey of related works followed by Section 3, which deals with the description of the methodology. Right after, in Section 4, the results of the research are presented and, finally, the conclusions.

2 Related Works

[15] examines how firms can improve CRM capabilities such as marketing and sales through Social Media usage, and his results demonstrate that social media can amplify the positive results of customer engagement and consequently, firm performance. [16] investigated the usage of Facebook Comments (F-comments) to verify how far such it is influencing the organizational performance of micro and small companies.

Similarly, [17] sought to verify whether the adoption of Facebook has an influence on the company's performance. As a result, the authors agree that the use of Facebook platforms for business purposes helps companies performance internally and externally. In addition, external pressures such as competitors and consumer influence are characterized as reasons to boost relationships with the customers. [18] examine the role of Social CRM in Social Information Systems (SIS), based on the results of four

case studies. The authors' conclusions show that the systems have applications that are not restricted to strengthening the company-customer relationship but can also be used to promote business in order to increase its expansion capabilities.

In the same article, [15] covered the technologies in the marketing area that are adopted in small businesses. The results show that managers see them as an opportunity to develop their relationship with customers. The authors also present a model built from the research's insights that covers all the phases of collecting information, building content, and measuring results.

Then [16] makes a study based on quantitative variables, obtained through an online survey, and qualitative, obtained from an exploratory search in the state of the art. [14] points out results in relation to Social CRM based on case studies built from interviews, providing quantitative and qualitative data. However, the approach was made in Large Companies, leaving the reality in which this work is inserted.

3 Methodology

This study was conducted in a quantitative way, through the application of a survey. And qualitative, through case studies. After recognizing the literature and defining the relevant concepts, the scope of the work and the construction of the research questions were defined. The scope was limited to MNCs and Individual Micro-Entrepreneurs (MEI) who use Social Media in their businesses. The research questions, presented in the Introduction, were defined from the gaps in the existing literature. The quantitative area was collected through a self-administered online survey divided into three sub-sections: i) general information about the company (e.g. size, the activity segment, social media management, etc.); ii) social media management (e.g. media used, published content, difficulties encountered, etc.); and, iii) customer relationship management (e.g. perception of the CRM concept, software used, etc.). The response method was divided between every and Likert-scale (from 1 to 5). The response time could vary between 5 and 10 minutes.

The second phase occurred with the collection of qualitative data through the case study methodology adapted from [17]. Four MNCs were selected, who actually participated in the events promoted by our research group. To guarantee unbiased results, all four companies belong to different market niches, described in the subsequent Case Studies. The data collection was conducted through semi-structured interviews lasting between 30 and 60 minutes. Table 1 presents the script defined in conjunction with practitioners.

Table 1. Script for conducting semi-structured interviews

Topics	Questions
Company data	General information of the interviewee
General information about social media	What strategies Social Media? Do you have a strategy? If so, how does it happen? How do the activities take place when using the networks?
Advantages for using social media in business	What are the perceived benefits of adopting social media in business? Were external factors perceived as influencing the motivation for using social media?

Secondary and general evaluation	How do you define the success of a publication? Do you invest money in Social Media? Does investing in Social Media generate value or difference to the business? Do you adopt CRM and/or Social CRM strategies in the company? How do you incorporate Social CRM? What are the reasons for adoption? What is responsible for the adoption? How is social management done? Which social networks and services are used and used? What are the diffusion arguments?
"Social CRM CRM"	What are the strategies chosen to face the market? What are the predominant perspectives?

It is important to highlight that all interviews were conducted through telephone calls, with the purpose of avoiding social context, since this phase of the research coincided with the period of dissemination of the new coronavirus. The results were first organized individually in a feedback format for the interviewees, highlighting the positive and negative aspects of the business in the area of customer relations. After that, the information collected was cross-checked in order to assess the equality and differences between the companies' processes. The results were discussed in partnership with specialists in Social CRM and Digital Marketing. The research results were organized and summarized in this work.

4 Results

In this section, we present the results from the survey and interviews, summarizing and discussing some relevant findings.

4.1 Companies Characterization

This survey was a prerequisite to attend some training courses provided by our research group. We used this approach instead of randomizing the survey, because it was possible to co-validate the answers. Moreover, it provided a good engagement with the open questions. We obtained 31 responses from SMEs, DSE, and informal businesses that already use Social Media in their business. Table 1 shows the basic information of the companies.

Table 1: General information of the companies

Type	Total in numbers	Percentage
SMEs	3	9.7%
DSE	14	45.2%
Informal	14	45.1%
Operating segment	Total in numbers	Percentage
Foodservice	1	3.2%
Advisory and communication	1	3.2%
Entertainment	1	3.2%
Engineering and Information Systems	1	3.2%

Creating and maintaining	4	52.5%
Search and web-browsing	2	25.0%
Voluntary activities	1	12.5%

Table 3 shows that entrepreneurs consider the internet as a very important channel for the development of their businesses. For this question, the entrepreneurs gave their answers on a 5-item scale ranging from 1 to 5. For a better understanding of the result, groupings were made: between 1 and 2 were grouped as "low relevance", 3 and 4 as "medium relevance", and 5 was called "high relevance".

Table 3. Internet and social media importance

Internet relevance for business	Total in numbers	Percentage
Low relevance	1	12.5%
Medium relevance	2	25.0%
High relevance	17	87.5%
Who manages social media	Total in numbers	Percentage
Owner/Manager	19	95.0%
Employee	1	5.0%
Outsourced	1	5.0%

Also in Table 3, it is possible to verify that, in almost all cases, the owner or manager is responsible for managing the company's social networks. Regarding the use of social media, it is clear that the most widespread are WhatsApp with 17 positive responses, followed by Instagram with 14 and Facebook with 11 of the 71 companies. This result agrees with [21], which says that these platforms are adopted because they are free of charge and easy to access. This distribution can be explained by the fact that these three social media are among the most used by Brazilians, only behind YouTube [22]. From this information, it can be inferred that the use of YouTube for business is still undervalued by companies, as in this research it represents only 4.9% of the sample, which, in turn, corresponds to only 2 positive responses, thus identifying gaps in the adoption of social media by small businesses.

Table 4. Types of content published on social media

Types of content	Total in numbers	Percentage
Selling products	25	62.5%
Promotion	18	44.1%
Advertising	17	41.3%
Relaxation information for customers	15	36.6%
Feedback service	14	34.1%
Institutional statement of the company	13	31.7%
Creating and designing	7	17.1%

In Table 4, it is possible to see that the main purpose of using social media is for the dissemination of products and services closely associated to companies, followed

by the sharing processes and construction of advertisements for new products. However, one of the least used topics is the search for customer feedback, which corresponds to less than 10% of the companies. It is known that one of the purposes of using social media for business is to create proximity between consumer and company, therefore, instigating sharing of shopping and relationship experiences with the company is of considerable importance for the strengthening of relations of customer loyalty.

Regarding the difficulties encountered by managers in the use of online social networks, Figure 1 shows that the biggest complication is related to the lack of knowledge of tools for the management of these platforms. In addition, a recurring topic among entrepreneurs is also the lack of time devoted to creating content and evaluating results. This can be explained by the information present in Table 1, wherein 46.6% of the cases it is the owner or manager who is responsible for these tasks. It is known that entrepreneurs of small businesses can perform several functions within his company, which makes him, most of the time, overloaded with tasks. In this way, they do not have time to learn new tools, nor learn to deal with the reports produced by them, a fact that directly impacts the management of social media in business.

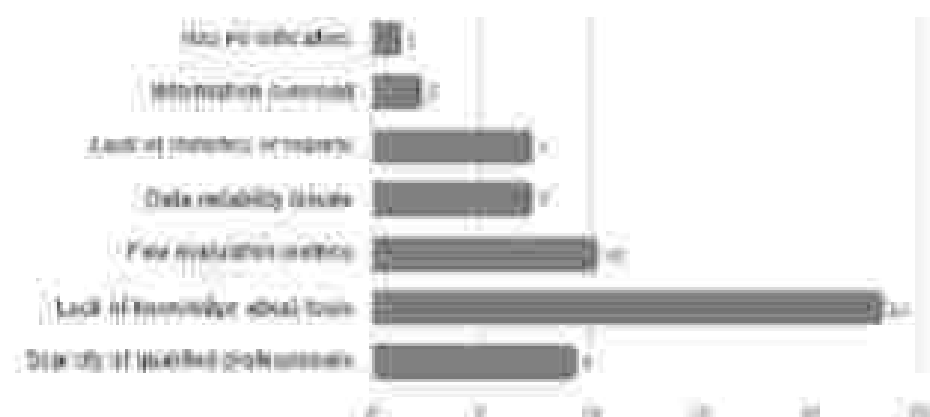


Fig. 1. Difficulties in managing results in social media.

Referring to Social CRM, Figure 2 says that 21 companies answered that they do not know how to conceptualize it. From the 3 that give their opinion, only 2 said that it is present in the entire process of relationship with customers. The other 4 responses were related only to the sales aspect. These examples can be found in Figure 2.



Fig. 1. Conceptualization of Social CRM by small businesses

4.1. Case studies

To conduct the case studies, four companies from different sectors were selected from the 11 companies that attended our surveys. It is important to justify the selection of the names of the companies, as this was one of the criteria agreed upon between the researchers and the interviewees, namely: **Retail Company** - belongs to the retail footwear and clothing industry. It is operation since 2009. **Food Company** - operates in the sector of supplying prepared foods, selling beverages, and the like. It has been on the market since 2013. **FET Company** - offering services for small animals, such as veterinary care and beautification. It has been in operation since 2013. **Photography Company** - its activities are described as a filming service organization. It has been offering its services since 2013. This information was collected on the Federal Government Register's free access platform, where it is possible to search in the National Register of Legal Entities using the companies' fancy name. All participants (interviewees) were the owners of the companies. More information about the profile of each interviewee can be seen in Table 3.

Table 3. General information of the interviewees

	Retail	Food	FET	Photography
Region	Alagoas	Paraná	Paraná	Paraná
Professional experience	Worked in a shoe store, in a retail store company, and owner of some appliances	No previous work experience	No previous work experience	Worked in a hair styling salon for five years
Responsibility	Work-related	Owner, styling and styling (Gestalt Theory)	Owner, styling and styling (Gestalt Theory)	Owner, styling and styling (Gestalt Theory)

In the case of the Food, Photography, and FET companies, the interviewees reported that they themselves are responsible for managing the business's digital platform. The company Retail faced two phases. In the beginning, the owner was the one responsible for taking care of social networks, while currently, a third-party company manages their social media channels. Only the Food and Retail companies have participated in

external consultancies offered by the Brazilian Micro and Small Business Support Service known by its acronym SEBRAE.

Then, it was moved to the questioning stage about the information that the owners have in relation to the use of social media. The responses were very similar, focusing on "brand presence", "product promotion" and "getting the customer's attention". A relevant fact was that some of the three companies cited social media as a channel of communication with customers. Therefore, a gap is identified here in the use of these platforms for business. With regard to the external factors that lead to the adoption of these platforms, the companies Pettil and Food cited the competitor's presence as a direct influence on the owners with the digital presence. PET talked about the collection it receives from customers in relation to the facilities of procedures that are done at the clinic, how animals are treated, etc. Photography reported that it uses social media to evaluate competitors.

In sequence, the most used metric by companies was the number of likes that a publication receives. Only Pettil cited the number of people interested in a particular product that was posted on social media, demonstrating that social media is being used for brand identification, which guides the portfolio selection process. After that, it was entered in the section regarding Social CRM and its processes. First, respondents were asked whether they knew how to conceptualize Social CRM. None of the four could answer. This fact was evident when the companies Pettil and PET said that they have software responsible for cash flow, inventory control, customer database, and sales control. The Food company also said it uses software to control orders. The company Photography reported that it used measuring customer satisfaction that participating in a course related to the topic. In other words, companies have adopted CRM and Social CRM processes, but they still do not understand what it is and where it is located.

Regarding the difficulties experienced, the Food portfolio selection process spoke about the limitations of the physical space, which is still small to serve the public it receives. Another point was the administration of multiple virtual service channels. Orders arrive via WhatsApp, phone, Instagram/Facebook messages, and ordering application - in addition to on-site customers. One of the solutions was to transcribe orders, excluding calls and messages via Instagram and Facebook. Despite dealing with many customers, there was an improvement in the speed of service and a reduction in any delays and errors in orders, increasing the level of satisfaction of service users.

In the PET company, the biggest problems are related to the time management of the owner, who sometimes needs to perform many functions and ends up being overwhelmed with work. An interesting dynamic in the relationship with the client was the use of WhatsApp as a tool for bringing together tutors and foster pets - daily, the manager sends photos and videos of the pets (posted in a directory) to each tutor. Some of the photos were posted on Instagram, but this action was canceled, given that tutors who did not have their pets microchipped complained to the manager. It should be noted that the manager stated that she does not feel comfortable using Facebook, which is why the company abandoned this platform. Photography cited sensitivity as a problem faced by the company. In periods of low demand, one solution was to produce diversification - such as photography courses, to increase its promotions, and the use of sponsored posts to increase the visibility of these actions.

Finally, questions were asked in the closing regarding the COVID-19 pandemic period. The PET company implemented the "Don't Log" service in order to reduce the

impact caused by the decrease in the circulation of people and restrictions imposed by social isolation. The company Retail is launching its online sales through Facebook, Instagram, and WhatsApp, intensifying the campaigns carried out on the latter platform. The Food company stated that it will continue with the delivery service and that it will also take the time without moving to the physical space to start the restaurant and expand the environment to be able to accommodate customers with social distancing. Photography said it was necessary to postpone the work and, currently, they are betting on the gastronomic photography market for restaurant delivery services.

After grouping the facts and information obtained through the survey and case studies, individual reports were built for the four companies that participated in the interviews. For these reports, aspects such as feasibility and desirability were taken into account in carrying out the suggested tasks. These tasks were grouped into positive points and points that could be improved with regard to the adoption of social media for business. Table 4 shows the information contained in the reports, which were also presented to the research partners in order to validate these suggestions.

Table 4. Social media platform reports

	Positive points	Points to improve
Retail	Interactions in "Buying publications", connecting to commercial service for managing social media, using WhatsApp Business as a communication channel, using software for managing internal processes	Automation of responses in WhatsApp Business, interaction with the public in social media through stories, creation of the company's social media presence or social presence
Food	Use of delivery service platform, interaction with the public in comments and reports, and active presence on social media	Definition of digital presence and branding, study ways to adapt WhatsApp Business as a communication channel, company registration on Google My Business, adoption of software for publishing advertisements
PEP	The stronger control between posts, use of WhatsApp Business as a communication channel, search for customer feedback, customer interaction in stories	Automation of responses in WhatsApp Business, use of social media for market research, not leaving all social media of the company, consideration of service
Photography	Adaptation of the business to live with market needs, interactions in "Buying publications", research in optimizing the feedback process with customers, and customer relation	Studying ways to adapt WhatsApp Business as a communication channel with customers, building a database with customer information, adopting appropriate business strategies

3 Lessons learned and perspectives

From the results, it is possible to see that companies will make use of a few functionalities that are made available within social media, for that of Customer Relationship Management. For example, the metrics used by entrepreneurs are mainly the number of likes, which is just one of the data offered by the Facebook and Instagram platforms to make the success of a publication. If they also had access to the engagement data of the publications and reach of people, they could more broadly identify what content is being well received by the public. With this, they would be able to build more targeted and effective campaigns and, consequently, save efforts in time and money, which are two of the main pains reported by managers.

A point that is also common among the four companies studied is the use of WhatsApp Business as a business tool. This platform can assist them in several ways, such as automating frequent responses, in which the manager saves time in the first contact with potential customers. In addition, WhatsApp Business can serve as a direct communication channel with customers to disseminate news and promotions and act as a catalog of products and services, in order to optimize the work of the social media manager.

Another aspect that was observed has a direct relationship with the branding of companies. According to [15, 14], branding can be defined as the set of strategies that define the brand. It is everything that involves it: from colors and shapes to define the value of products and services and sales practices. Moreover, for small companies, this can make a huge difference, since it serves to position the brand in the market and help to identify it among others. This process can be initiated with the definition of the values perceived in the products, logo creation, and colors and visual elements definition, which will identify the company. After these steps, the visual identity creation process can begin, in which the way in which the company will communicate with its public will be defined. Then, we move on to the digital presence stage, where the company will set up its social media channels.

These channels are defined according to the company's needs. For example, the food company cited Food as one of its sales channels. In addition, for the Retail company, the report sent to the manager indicated the configuration of the store's marketplace, where he will be able to count on another online sales channel, in addition to the presence. This module is known for bringing together in a single place, several offers of products and services from different companies, such as an online catalog (e.g. Amazon, Mercado Livre, OLX, etc.). Facebook currently has its own marketplace too. The correct definition of social media can make a big difference in the sales volume achieved by companies.

To assist managers in these adjustments, it is therefore necessary to offer individualized courses and consultations, with the purpose of putting them in the adjustments that were suggested in the reports. It is also important to assist them in understanding metrics for evaluating results and how to optimize their internal processes. With this, it is expected to positively impact its business in the growth of brand visibility and volume of sales.

6 Conclusions

This article analyzed quantitative aspects, through a survey, and qualitative aspects, with the help of case studies, of the implementation of Social CRM in Micro and Small Companies, a sector that is highly relevant for Brazilian trade. The purpose of these analyses was to identify which aspects of Social CRM are relevant to SMEs (RQ1) and how they implement them in their business (RQ2).

The results referring to RQ1 show that entrepreneurs classify the Internet as a highly relevant means of communication for their businesses, with Facebook, Instagram, and WhatsApp as the main platforms. They use these media to publicize their products and promotions, in addition to disseminating news and promoting investigations. However, only a portion of managers uses them as a channel for seeking feedback from their customers, which ends up neglecting the loyalty process. The results also show that entrepreneurs are unaware of Social CRM concepts and applications. Additionally, the lack of this knowledge means that they are unable to achieve better results since they themselves are responsible for managing social media. As a solution to this problem, short and asynchronous training on this topic can be offered to managers.

Regarding RQ2, the results obtained indicate that the main reason for using social media is directly linked to the presence of the brand. However, as in the previous phase, it became evident that managers will do not use social networks as a two-way communication channel, which is used only for the broadcasting of products and services. A fact that drew attention was that entrepreneurs implement CRM processes in their companies, such as using software for cost and inventory management; at the same time, they report that they do not know the concept, which also contributes to the neglect of related functionalities. Despite this, the predominant forced micro and small companies to adapt to the new reality, demonstrating their adaptability with actions that include launching new service packages and promotions, in addition to offering new services through online sales and delivery.

In view of the results, this work contributes to the visualization of the main gaps in the Customer Relationship Management sector in social media, in companies among small companies. We concluded that Social CRM is still an unexplored environment for micro and small companies, but with great potential to boost sales, develop customer loyalty and increase brand visibility. The lessons learned have the potential of guiding policymakers to create more adequate measures for boosting the market sector. As future work, it is intended to plan a set of actions with an international team, in order to help entrepreneurs to explore the functionalities of social media, thus contributing to the optimization of their internal and external processes.

7 Acknowledgment

We would like to thank the Federal University of Western Pará and the National Council for Scientific and Technological Development (CNPq) for their financial support for conducting this work, and the research partners from the State University of Maricá for their direct collaboration in conducting the research. We would also

17. 23 (2017) <https://doi.org/10.17718/ma.23.17-03>.
18. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840. 841. 842. 843. 844. 845. 846. 847. 848. 849. 850. 851. 852. 853. 854. 855. 856. 857. 858. 859. 860. 861. 862. 863. 864. 865. 866. 867. 868. 869. 870. 871. 872. 873. 874. 875. 876. 877. 878. 879. 880. 881. 882. 883. 884. 885. 886. 887. 888. 889. 890. 891. 892. 893. 894. 895. 896. 897. 898. 899. 900. 901. 902. 903. 904. 905. 906. 907. 908. 909. 910. 911. 912. 913. 914. 915. 916. 917. 918. 919. 920. 921. 922. 923. 924. 925. 926. 927. 928. 929. 930. 931. 932. 933. 934. 935. 936. 937. 938. 939. 940. 941. 942. 943. 944. 945. 946. 947. 948. 949. 950. 951. 952. 953. 954. 955. 956. 957. 958. 959. 960. 961. 962. 963. 964. 965. 966. 967. 968. 969. 970. 971. 972. 973. 974. 975. 976. 977. 978. 979. 980. 981. 982. 983. 984. 985. 986. 987. 988. 989. 990. 991. 992. 993. 994. 995. 996. 997. 998. 999. 1000.