

Análise do setor de telecomunicação brasileiro: Uma visão sobre Reclamações

Gustavo Nogueira de Sousa¹, Isabelle da Silva Guimarães², Julio Augusto Nogueira Viana³, Olaf Reinhold³, Antonio Fernando Lavareda Jacob Junior¹, Fábio Manoel França Lobato^{1,2}

sougusta@gmail.com, isabelle.guimaraes2@gmail.com, julio.viana@scrc-leipzig.de,
reinhold@wifa.uni-leipzig.de, antonio.jacob@gmail.com, fabio.lobato@ufopa.edu.br.

¹ Universidade Estadual do Maranhão (UEMA), São Luís, Brasil

² Universidade Federal do Oeste do Pará (UFOPA), Santarém, Brasil

³ Social CRM Research Center (SCRC), Leipzig, Alemanha

DOI: [10.17013/risti.37.31-48](https://doi.org/10.17013/risti.37.31-48)

Resumo: Mídias digitais estão cada vez mais presentes no cotidiano do ser humano. Este fato contribui para que o volume de conteúdo gerado por usuário aumente consideravelmente. De um ponto de vista prático, as análises desses dados requerem diferentes perspectivas e métodos para se obter resultados satisfatórios. Essas análises podem subsidiar a tomada de decisão por gestores por meio da identificação de necessidades e problemas, guiando o processo de melhoria continuada 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 orientem a tomada de decisões 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 provisão de uma abordagem que auxilia no planejamento estratégico de corporações, levando em consideração situações reportadas pelos consumidores.

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

Analysis of the Brazilian telecommunications sector: An overview on Complaints

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 analyzes 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 relationship. 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 consumers.

Keywords: Social Media, Text Mining, Customer Relationship Management, Social CRM, Complaints.

1. Introdução

As mídias sociais oferecem uma ampla gama de funcionalidades que permitem ao usuário o compartilhamento e o consumo de conteúdo online (Carr & Hayes, 2015; Silva et al., 2018). Este fenômeno impacta diretamente no relacionamento entre empresas e consumidores. Isso ocorre devido à popularização da internet e o aumento no acesso a essas plataformas sociais que, consequentemente, tornam os consumidores mais engajados com marcas e na troca de informações sobre produtos e serviços (Lobato et al., 2017). Consequentemente, percebe-se um aumento substancial na quantidade de conteúdos gerados pelos usuários (User Generated Content – UGC) (Bahtar & Muda, 2016; Lobato et al., 2017; Nusair, Hua, Ozturk, & Butt, 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, Lobato, & Cirqueira, 2017; Schmäh, Wilke, & Rossmann, 2017). Este cenário impõe em 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 produto e serviço (Alt & Reinhold, 2012).

eWoM é visto como um forte determinante na decisão de compra, haja vista que cerca de dois terços dos consumidores verificam as avaliações de produto, serviços e marcas antes de decidirem adquiri-los (Ahmad & Laroche, 2017; Constantinides & Holleschovsky, 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 a serem resolvidos (Gavilanes, Flatten, & Brettel, 2018; Einwiller & Steilen, 2015).

Pesquisas na literatura revelaram que há poucos trabalhos que realizam a análise e extração de conhecimento de eWoM expressas em plataformas de reclamações online. Encontra-se o uso massificado de diversas plataformas de eWoM como fonte de dados para a geração de conhecimento, tais como: Facebook (Bahtar & Muda, 2016; Kim & Johnson, 2016; Liu, Li, Ji, North, & Yang, 2017; Vermeer, Araujo, Bernritter, & van Noort, 2019), Twitter (Chakraborty et al., 2017; Einwiller & Steilen, 2015; Vermeer et al., 2019) e reviews em loja de aplicativo (Ali, Joorabchi, & Mesbah, 2017; McIlroy, Ali, Khalid, & E. Hassan, 2016; Vu, Nguyen, Pham, & Nguyen, 2016).

No Brasil, plataformas de reclamações online têm bastante relevância e influência. Segundo o site (Alexa, 2019), no período de 01/11/2019 a 01/02/2020 o *ReclameAqui*¹ esteve entre os 25 sites mais acessados do país, no qual, diariamente cada usuário fez em média três visitas na página com um tempo médio de três minutos nas interações. Além disto, percebeu-se que as empresas de telecomunicações são as mais mal avaliadas de acordo com o *ranking* que considera as 120.000 cadastradas na plataforma². 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 nortearam 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 geo-temporais?
- 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, igualdade de oportunidade, facilidade de transações e comunicação entre indivíduos (Bankole, Osei-Bryson, & Brown, 2015; Mujahid, Sierra, Abdalkareem, Shihab, & Shang, 2018; Sharma, Fantin, Prabhu, Guan, & Dattakumar, 2014). Devido às características deste trabalho, as empresas foram selecionadas considerando o número de clientes e a presença em todo o território nacional, e de acordo com a (Telecomunicações, 2020), o *Market share* das empresas escolhidas neste trabalho representam juntas 97% de telefonia móvel, 72% da banda larga fixa, 96,9% de TV por assinatura e 94,4% 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.

2. Trabalhos Relacionados

Para que o conteúdo de mídia social seja útil para geração de conhecimentos que embasam a tomada de decisões, é necessário a definição e o uso de estratégias para a extração dos dados. Considerando que alguns dados são de difícil acesso, (Olmedilla, 2016) define uma arquitetura para um framework com diretrizes e abordagens a serem seguidas para a extração de dados. Com isto, mostrou-se como um *Webcrawler* pode ser extremamente eficaz no processo de reunir e identificar grandes quantidades de conteúdo gerado pelo usuário. Além disso, mostram a importância das ciências sociais e

¹ <https://www.reclameaqui.com.br/>

² <https://www.reclameaqui.com.br/ranking/>

da computação no processo de análise de dados sociais. Estudos semelhantes baseiam-se neste *framework*, a citar (D'Aquino Netto et al., 2019; Rodrigues et al., 2019).

Devido a informalidade dos textos 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 dos resultados. (Cirqueira et al., 2018) realizaram uma análise na literatura para identificar os principais métodos utilizados para tratar conteúdos escritos em Português-Brasileiro. Foram reunidos um total de 62 artigos relevantes, os quais possibilitaram a listagem dos principais métodos e etapas necessárias.

Em relação aos tipos de análises, a modelagem de tópicos merece destaque, uma vez que permite a identificação de certos padrões nos dados, nos quais seriam difíceis a descoberta manual. Para isso, (Ernala et al., 2018) utilizaram técnicas de modelagem 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 determinaram que menções, apoio emocional e discussões em torno da vida pessoal são fortes preditores de um ambiente em que é possível a divulgação dos mais variados tipos. Outros trabalhos apresentam a modelagem de tópicos como uma forma de extrair os termos importantes no texto, tal como (Aldous et al., 2019; Cirqueira, Pinheiro, et al., 2017; Almeida et al., 2017; Li et al., 2019).

Sob o ponto de vista da análise de eWoM, (Rhee & Yang, 2015) apresenta uma análise do setor de turismo a partir de *reviews* publicados em plataformas online. Os autores avaliaram seis aspectos relevantes em hotéis, com cinco tipos de viagens pré-definidos e viajantes de dois países diferentes. Como resultado, os autores verificaram que o “valor” e o tipo do “quarto” são os elementos que mais influenciam em uma avaliação. Assim sendo, a análise de conteúdos de mídias sociais permite que as ações nestas plataformas sejam mais efetivas e eficientes (Chirumalla, Oghazi, & Parida, 2018).

No trabalho de (Tang & Guo, 2013) destacam a viabilidade e eficácia do uso de técnicas de mineração de texto para processar o conteúdo. No mesmo sentido, (Carrascal, Cotte, Arango, & Vélez, 2019) apresenta a técnica de mineração de texto como forma eficaz para a descoberta de conhecimento e conteúdos textuais. Os autores realizaram a identificação das palavras mais relevantes em conjuntos textuais e com potencial de realizar provisões futuras com base nos resultados.

3. Metodologia

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 (Chinchilla & Ferreira, 2016). Considerando os níveis de abstração propostos, os trabalhos de (Rollins, 2015; Schafer, Zeiselmaier, Becker, & Otten, 2019; Wirth, 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 fases, conforme pode ser observado na Figura 1.

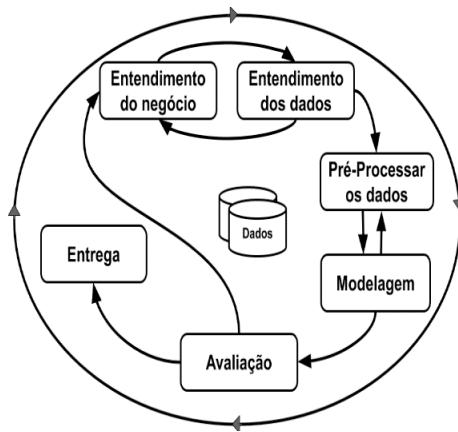


Figura 1 – Diagrama de funcionamento do modelo CRISP-DM (Adaptada de Wirth, 2000).

Devido a característica cíclica deste método, o processo de mineração de dados não é finalizado quando uma solução é implementada. Neste 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*.

3.1. Entendimento do Negócio

Tomando em consideração a maleabilidade que a metodologia CRISP-DM possibilita, nesta etapa é realizado o entendimento do contexto em que as análises poderiam ser aplicadas, a partir da definição dos objetivos da mineração de dados. Neste contexto, verificou-se que as reclamações podem representar um importante meio para obtenção de uma grande quantidade de informações autênticas sobre produtos e serviços feito de forma voluntária. Com base nisto, 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-temporais;
- Identificar as implicações práticas das reclamações nas empresas.

3.2. Entendimento dos dados

Nesta etapa são realizadas a coleta, descrição, exploração e verificação da qualidade dos dados coletados. Devido as dificuldades de utilização da API fornecida, os dados foram extraídos da plataforma de reclamações *online* chamada *ReclameAqui* por meio de um *WebCrawler* escrito em Python, o qual utiliza a biblioteca *requests*³. Esta plataforma

³ <https://requests.readthedocs.io/en/master/>

foi escolhida devido sua popularidade. Dentre todos os sites acessados no Brasil, o ReclameAqui é o 25º site com maior número de acessos (Alexa, 2019), sendo o site mais popular na categoria de reclamações.

Como alvo da extração, foram selecionadas as quatro maiores empresas do setor de telecomunicações brasileiro. 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* fornecido pelo ReclameAqui, o qual contém mais de 120.000 empresas cadastradas. Cada reclamação extraída era composta pelos dados detalhados na Tabela 1.

| Campos | Descrição |
|----------------------|---|
| Empresa ID | Identificação da empresa relacionada na reclamação |
| Reclamação ID | Identificação da reclamação |
| Título da Reclamação | Título dados pelo usuário a reclamação |
| Reclamação | Relato do problema reportado |
| Estado e cidade | Estado e cidade do consumidor que realizou a reclamação |
| Data/hora | Data e horário do relato |

Tabela 1 – 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 utilizado um conjunto amostral que 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.

3.3. 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 saudações, de URLs, *stopwords*, números, acentuação e de caracteres especiais (Cirqueira et al., 2018).

A remoção de saudações e de URLs significa que todo conjunto de caractere que representa uma saudação (e.g. “Olá”, “Oi”) ou um endereço de algum site (e.g. “www.site.com.br”) foi removido. Da mesma forma, números, acentuação e caracteres especiais foram retirados, visto que são desnecessários para as análises. Palavras que são consideradas *stopwords*, ou seja, palavras que não contribuem para o significado do texto (e.g. “e”, “de”, “em”) foram eliminadas.

3.4. Modelagem

Esta etapa foi dividida em duas fases: a aplicação da extração de tópicos; e a correlação entre os tópicos encontrados. Na primeira fase da 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 avaliação qualitativa e anotação dos tópicos percebeu-se que os termos obtidos pelo NMF estavam mais relacionados entre si e que representavam tópicos mais coerentes e diversos (Chen et al., 2019). Devido a isto, o *Term Frequency-Inverse Document Frequency (TF-IDF)* juntamente com o *NMF* foram adotados neste trabalho com o objetivo de classificar todas as palavras por ordem de importância no conjunto de textos, tal como demonstrado por (Salminen et al., 2018; Trstenjak, Mikac, & Donko, 2014).

A modelagem e análise de todas as reclamações foi realizada de acordo com a empresa sob uma perspectiva nacional e regional. Os tópicos foram determinados 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 coocorrência.

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

$$r = \frac{n!}{p!(n-p)!} \quad (1)$$

Sendo que r é a quantidade de registros resultantes; n é o número de palavras-chave do trabalho; 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)*, o qual que pode ser visualizado com o suporte de uma ferramenta de análise de redes. No presente estudo o *software Gephi*⁴ foi utilizado para este fim.

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 tópicos em termos nacionais e regionais é mostrada na segunda subseção. E, por fim, na terceira subseção são apresentadas as análises geotemporal das reclamações obtidas.

4.1. Extração dos dados

O processo de coleta dos dados resultou em um total de 397.950 reclamações. Considerando a dinamicidade do setor de telecomunicação e sua propensão a mudanças (Stone, 2015), nestas análises foram utilizadas as reclamações do período de 09/09/2018

⁴ <https://gephi.org/>

a 09/09/2019, resultando em 225.593 reclamações. Na Tabela 2 é apresentado as quantidades de reclamações coletadas em nível regional e nacional de cada empresa.

| | Norte | Nordeste | Centro-Oeste | Sudeste | Sul | Total |
|-------|--------------|-----------------|---------------------|----------------|------------|--------------|
| Tim | 1.377 | 10.100 | 3.176 | 47.339 | 9.282 | 72.174 |
| Vivo | 1.114 | 6.836 | 4.610 | 59.678 | 9.304 | 81.542 |
| Oi | 1.257 | 6.401 | 2.765 | 19.459 | 4.779 | 34.661 |
| Claro | 974 | 3.601 | 2.753 | 27.056 | 3.732 | 38.116 |

Tabela 2 – Distribuição dos dados pelas regiões do país.

O Brasil possui aproximadamente 210 milhões de habitantes distribuídos em 5 regiões (IBGE, 2019a). As reclamações analisadas foram estratificadas de acordo com a operadora e a respectiva região de origem, conforme pode ser observado na Tabela 2. A região Sudeste tem um destaque frente as demais. Esta região possuí aproximadamente 42% da população brasileira e apresenta uma taxa superior a 65% do total de reclamações coletadas. Este fato pode ser justificado, uma vez que a região sudeste é a mais rica 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, 2020).

4.2. Modelagem 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 descritos a seguir.

4.2.1. Panorama Nacional

O total de reclamações de cada empresa (apresentados na Tabela 2) foram utilizados para compor os tópicos de reclamação no panorama nacional. A partir da limitação do número de tópicos frente as categorias de problemas da plataforma, obteve-se a seguinte distribuição por empresa: 31 tópicos da Tim; 29 na Vivo; 26 tópicos na Oi; e 26 na Claro. Após uma análise inicial, alguns tópicos semelhantes (termos sinônimos, por exemplo) foram combinados. Além disso, foram considerados apenas tópicos únicos e relevantes. O resultado encontra-se sintetizado na Tabela 3.

Na Tabela 4 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. Estes 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 (Gavilanes, Flatten, & Brettel, 2018; Einwiller & Steilen, 2015).

| | Tópicos encontrados | Tópicos resultante |
|-------|----------------------------|---------------------------|
| Tim | 31 | 17 |
| Vivo | 29 | 19 |
| Oi | 26 | 17 |
| Claro | 26 | 19 |

Tabela 3 – Número de tópicos encontrados e resultantes

Para aprimorar a discussão dos resultados, alguns detalhes sobre o conhecimento do domínio fazem-se necessário, isto é, características inerentes às empresas analisadas. Por exemplo, a Tim oferece um plano exclusivo para seus clientes, o qual é acessível somente por meio de um convite enviado por outros usuários desse plano⁵. A partir da identificação do tópico relacionado é possível verificar que reclamações relacionadas a este serviço são frequentes e podem ser relacionados este processo de negócio utilizado pela empresa.

| Tim | Vivo | Oi | Claro |
|------------------------|------------------------|------------------------|------------------------|
| Atendimento | Atendimento | Atendimento | Atendimento |
| Ativação chip | Ativação chip | Cancelar serviço | Ativação linha |
| Cancelamento | Cancelamento | Cancelamento | Cancelamento |
| Cobertura móvel | Cobertura móvel | Cobrança indevida | Cobertura móvel |
| Cobrança indevida | Cobrança indevida | Instalação de | Cobrança indevida |
| Convite para plano | Internet residencial | equipamento | Internet móvel |
| Franquia de dados | Ligações telemarketing | Internet móvel | Ligações telemarketing |
| Ligações telemarketing | Linha cancelada | Ligações telemarketing | Linha cancelada |
| Loja física | Loja física | Linha cancelada | Loja física |
| Nome negativado | Mudança de endereço | Loja física | Multa por fidelidade |
| Linha cancelada | Mudança de plano | Multa por fidelidade | Nome negativado |
| Pagamento | Multa por fidelidade | Nome negativado | Número cancelado |
| Plano | Nome negativado | Plano controle | Pacotes de dados |
| Plano Controle | Número cancelado | Planos | Pagamento |
| Portabilidade | Pagamento | Portabilidade | Planos |
| Recarga | Portabilidade | Qualidade do serviço | Portabilidade |
| Valor serviço | Qualidade do serviço | Recarga | Recarga |
| | Recarga | Telefone fixo | Senha de acesso |
| | Visita técnica | | Serviço de música |

Tabela 4 – Modelagem de tópicos para avaliação do panorama nacional.

Para aprimorar a discussão dos resultados, alguns detalhes sobre o conhecimento do domínio fazem-se necessário, isto é, características inerentes às empresas analisadas. De forma geral, as reclamações encontram-se relacionadas a mais de um tópico. A correlação entre os tópicos das reclamações dos consumidores propicia a verificação da distribuição desta cadeia de insatisfação. Na Figura 2, são expostas as relações entre os

⁵ https://www.timbeta.com.br/timbeta/como_ser_beta

tópicos, sendo que tanto o tamanho das palavras quanto a espessuras e variação dos tons das arestas representam a importância e o peso da relação.

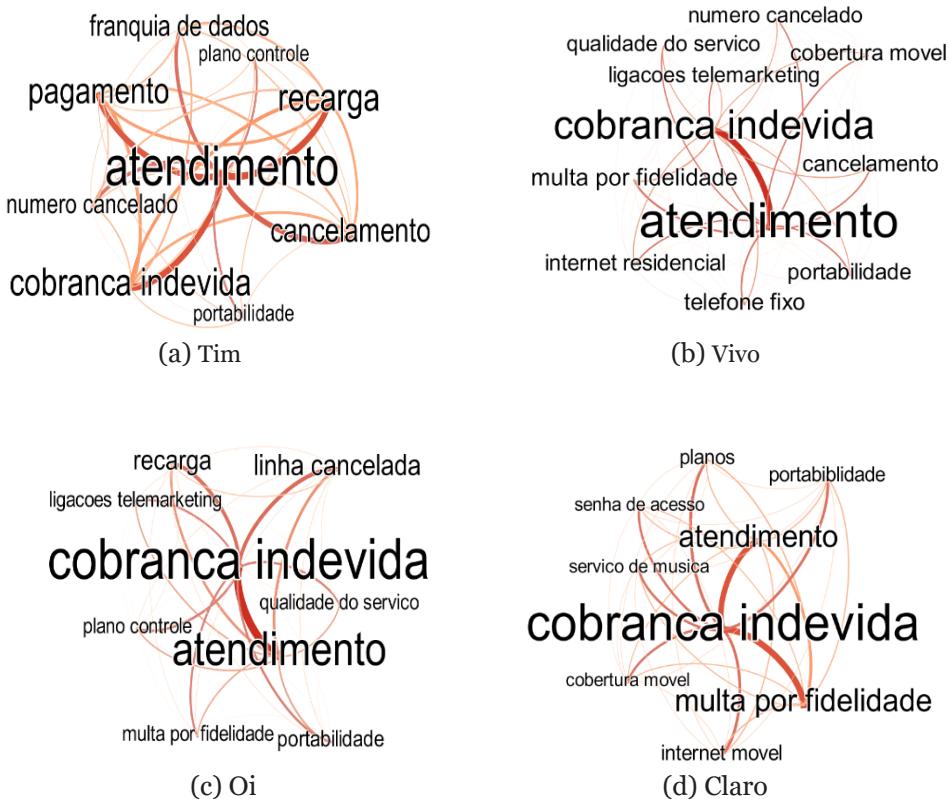


Figura 2 – Relação entre tópicos das reclamações

Ao se analisar os dados dispostos Figura 2, três pontos merecem destaque em relação às companhias estudadas:

1. A Tim, tem como seu principal problema o atendimento. Na Figura 2.a 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;
2. A Vivo e Oi apresentam como seus principais gargalos o atendimento e a cobrança indevida. Na Figura 2.b (Vivo) e na Figura 2.c (Oi) há uma grande associação entre os dois problemas, sendo que para a empresa Vivo o atendimento tem peso maior. Já para a empresa Oi, cobranças indevidas o fruto de maior frustração entre clientes. Isto indica que os processos de reclamação sobre

cobrança indevida não são tratados satisfatoriamente pelos canais de, gerando descontentamento para seus clientes;

3. A Claro tem um conjunto de 3 tópicos fortemente associados, que são a “cobrança indevida”, “atendimento” e “multa por fidelidade”, como observado na Figura 2.d. Essa associação de tópicos indica uma forte relação entre os três assuntos e facilita a análise do teor da reclamação.

4.2.2. Análise regional

A modelagem também foi realizada considerando uma perspectiva regional. Os tópicos obtidos para cada região são apresentados na Tabela 5. Os tópicos que iniciam com “+” são específicos para a empresa no contexto regional. Já os que iniciam com “-” são tópicos que não faz parte do contexto da região relacionada.

| | Norte | Sul | Sudeste | Nordeste | Centro-Oeste |
|------|--|--|--|---|---|
| Tim | + acesso a serviços + multa por fidelidade + renovação de pacote + serviço de música | + internet móvel + promoção + sms indevido | + acréscimo nos valores + multa por fidelidade + serviço de música | + mudança de plano + multa por fidelidade + promoção + serviço de música + sms indevido | + internet móvel + multa por fidelidade + promoção + qualidade do serviço + serviço de música |
| Vivo | - plano controle | - cobertura móvel - plano controle | - cancelamento - plano controle | - loja física - número cancelado - plano controle | - cancelamento - plano controle |
| | + celular bloqueado + instalação de equipamento + plano + sms ilimitado | | + planos | + entrega de compras | + cancelamento |
| | - internet residencial - ligações telemarketing - linha cancelada - mudança de endereço - multa por fidelidade - qualidade da internet - qualidade do serviço - telefone fixo - visita técnica | - ativação chip - cancelamento | - cancelamento - qualidade da internet - telefone fixo | - mudança de endereço - qualidade da internet - relação - telefone fixo | - internet residencial - mudança residencial - telefone fixo |

| | + contrato + pagamento | + manutenção + renegociação | + manutenção | + pagamento + renegociação | + renegociação |
|-------|--|---|--|---|---|
| Oi | - multa por fidelidade | - cancelamento - plano controle | - plano controle - telefone fixo | - cancelamento - cobertura móvel - ligações - telemarketing - plano controle - qualidade do serviço - telefone fixo | - cancelamento - plano controle - qualidade do serviço - telefone fixo |
| | - plano controle - telefone fixo | | | | |
| Claro | - cancelamento - linha cancelada - loja física | - linha cancelada - loja física | - número cancelado - pacotes de dados | - cancelamento - linha cancelada - pacotes de dados - serviço de música | - linha cancelada - loja física - pacotes de dados - recarga - senha de acesso - serviço de música |
| | - pacotes de dados - serviço de música | - pacotes de dados - serviço de música | | | |

Tabela 5 – Tópicos por regiões das empresas estudadas.

Contrastando-se as Tabelas 4 e 5, cenário nacional e regional, respectivamente, é possível evidenciar as particularidades de cada empresa por região. Por exemplo, na Tim é possível notar que as regiões com mais problemas que diferem do contexto nacional são a Nordeste e Centro-Oeste. Para a empresa Vivo, a região Norte apresenta a maior diferença em relação aos problemas nacionais. Nela não há oito tópicos de problemas nacionais e há quatro novos tipos de problemas que são específicos da região.

Para a empresa Oi, o tópico “renegociação” é um problema específico das regiões Sul, 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 tópicos de problema o “número cancelado” e o “pacote de dados”.

Diante disso, empresas de telefonia podem analisar quais os tópicos mais reclamados em uma determinada região, bem como os tópicos que não são relevantes (no período estudado) para os clientes em cada área do país. Tal informação permite que áreas menos populosas e com menos reclamações sejam tratadas de acordo com a sua especificidade, eliminando o possível viés que as áreas mais populosas podem causar em análises que consideram reclamações de todo país e elencam somente problemas nacionais.

É importante ressaltar que, a análise de relação entre tópicos de reclamação exemplificado na subseção 4.2.1, é perfeitamente aplicável nos contextos regionais, estaduais e municipais, no entanto por questões de espaço estas análises foram suprimidas do presente trabalho.

4.3. Distribuição “Geo-Temporal”

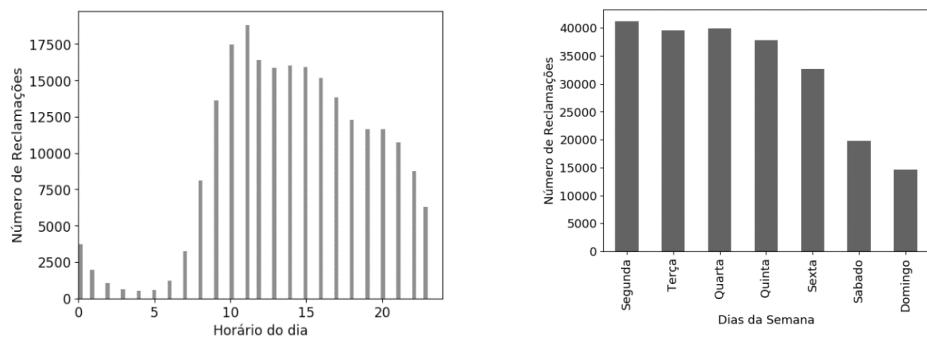
As reclamações coletadas estão distribuídas em todo território do Brasil. São distribuídas de forma heterogênea e desproporcional a taxa populacional de cada região. Na Tabela 6 é possível observar como se dá a distribuição geográfica das reclamações no país através

da comparação das taxas populacionais, PIB, Linhas ativas e de reclamações em cada região. Os dados são apresentados de forma a relacionada ao PIB, taxa populacional e de linhas móveis ativas em cada região.

| | População | PIB | Linhas ativas | Reclamações |
|--------------|------------------|------------|----------------------|--------------------|
| Norte | 9% | 6% | 7% | 2% |
| Nordeste | 27% | 10% | 24% | 12% |
| Centro-Oeste | 14% | 17% | 8% | 6% |
| Sudeste | 42% | 53% | 46% | 68% |
| Sul | 14% | 14% | 15% | 12% |

Tabela 6 – Proporção populacional (IBGE, 2019), PIB (“Produto Interno Bruto - PIB | IBGE,” n.d.), Linhas ativas (“Móvel Pessoal - Files - ownCloud,” n.d.) e de reclamações por Região

Na Figura 3, são expostas as quantidades de reclamações por horário e dia da semana, sendo que na Figura 3a contém a distribuição de todas as reclamações em 24 horas e na Figura 3b contém a distribuição de todas as reclamações nos dias da semana.



(a) Hora em que foram realizadas as reclamações (b) Número de reclamações por dia da semana.

Figura 3 – Número de reclamações por dia da semana.

Observando-se a Figura 3 é possível perceber que não há diferenças significativas entre os meses do ano em relação ao número de reclamações. A distribuição em relação aos dias da semana e horários seguiram também estáveis com variações para baixo nos finais de semana e entre onze horas da noite e oito horas do dia seguinte. No entanto, essas informações podem auxiliar as empresas na definição de tarefas e processos internos para solucionar os problemas levantados nessa plataforma online de reclamações de forma ágil e eficiente.

A análise dos dados dispostos na Tabela 6 e figura 3 nos revela alguns *insights* importantes sobre a distribuição das reclamações no país, a saber:

- A região Norte é a região com menor PIB do país, menor número de habitantes e em quantidade de linhas de celulares ativas. Há uma certa discrepância em relação à proporção de reclamações de outras regiões. Todas as outras tiveram a taxa de reclamação mais próxima do número de linhas ativas;
- A região Centro-Oeste tem cerca de 8% do total de linhas ativas e 6% das reclamações coletadas são referentes a essa região;
- A 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 22 pontos percentuais acima do número de linhas ativas;
- Todas as outras regiões, exceto a região sudeste, tiveram taxas de reclamações inferiores a taxa de linhas ativas. O que pode indicar que os clientes da região Sudeste são mais propensos a utilizar plataformas de reclamações online para realizar suas queixas.

5. Considerações Finais

Neste artigo, foram analisadas reclamações de quatro empresas do setor de telecomunicações brasileiro objetivando determinar quais eram os principais tópicos nas reclamações, como elas se relacionam entre si e qual a sua distribuição geotemporal. Para tal, utilizou-se a metodologia CRISP-DM, sendo que na fase de extração de conhecimento foram aplicados métodos de modelagem de tópicos a nível nacional e regional. Os resultados mostram que o uso de técnicas de mineração de texto combinado com uma análise geo-temporal permite a antecipação e a correção de problemas, a análise de grandes volumes textuais de forma rápida e análises estratégicas do, 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 *PP1* – 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 *PP2*, - identificou-se que a relação entre os tópicos é diferente em cada empresa e indicam que a cadeia de problemas a serem solucionados variam entre os concorrentes. Acerca da *PP3*, verificou-se que existe uma diferença regional em relação aos tópicos das reclamações analisadas, porém não foram identificadas diferenças significativas no número de reclamações durante o ano. Por fim, por meio da resposta da *PP4*, os resultados auxiliam as empresas de telecomunicação 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 provisão de uma abordagem de análise de reclamações que identifique as reais necessidades dos usuários 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 tratadas 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, pois não incluem 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.

Referências

- Ahmad, S. N., & Laroche, M. (2017). Analyzing electronic word of mouth: A social commerce construct. *International Journal of Information Management*, 37(3), 202–213.
- Aldous, K. K., An, J., & Jansen, B. J. (2019). View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organizations. In Proceedings of the International AAAI Conference on Web and Social Media. (pp. 47–57).
- Alexa. (2019). Alexa - Top Sites in Brazil - Alexa. Retrieved February 3, 2020, from: <https://www.alexa.com/topsites/countries/BR>
- Ali, M., Joorabchi, M. E., & Mesbah, A. (2017). Same App, Different App Stores: A Comparative Study. In *Proceedings - 2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems, MOBILESoft 2017*, (pp. 79–90).
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). *A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques*. New York: arXiv e-prints.
- Almeida, G. R. T. de, Lobato, F., & Cirqueira, D. (2017). Improving Social CRM through eletronic word-of-mouth: a case study of ReclameAqui. In *XIV Workshop de Trabalhos de Iniciação Científic*.
- Alt, R., & Reinhold, O. (2012). Social Customer Relationship Management (Social CRM). *Business & Information Systems Engineering*, 4(5), 287–291.
- Bahtar, A. Z., & Muda, M. (2016). The Impact of User – Generated Content (UGC) on Product Reviews towards Online Purchasing – A Conceptual Framework. *Procedia Economics and Finance*, 37, 337–342.
- Bankole, F. O., Osei-Bryson, K. M., & Brown, I. (2015). The Impacts of Telecommunications Infrastructure and Institutional Quality on Trade Efficiency in Africa. *Information Technology for Development*, 21(1), 29–43.
- Carr, C. T., & Hayes, R. A. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication*, 23(1), 46–65.
- Carrascal, A. I. O., Cotte, D. S., Arango, N. A. R., & Vélez, A. F. P. (2019). Descubrimiento de Conocimiento en Historias Clínicas mediante Minería de Texto. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, (34), 29–43.

- Chakraborty, A., Messias, J., Benevenuto, F., Ghosh, S., Ganguly, N., & Gummadi, K. P. (2017). *Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations*. In *Proceedings of ICWSM*, 22–31.
- Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*, 163, 1–13.
- Chinchilla, L. D. C. C., & Ferreira, K. A. R. (2016). Analysis of the behavior of customers in the social networks using data mining techniques. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, (pp. 623–625).
- Chirumalla, K., Oghazi, P., & Parida, V. (2018). Social media engagement strategy: Investigation of marketing and R&D interfaces in manufacturing industry. *Industrial Marketing Management*, 74(February 2017), 138–149.
- Cirqueira, D., Fontes Pinheiro, M., Jacob, A., Lobato, F., & Santana, A. (2018). A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, (pp. 746–749). IEEE.
- Cirqueira, D., Pinheiro, M., Braga, T., Jacob, A., Reinhold, O., Alt, R., & Santana, Á. (2017). Improving relationship management in universities with sentiment analysis and topic modeling of social media channels. *Proceedings of the International Conference on Web Intelligence - WI '17*, (pp. 998–1005).
- Constantinides, E., & Holleschovsky, N. I. (2016). Impact of Online Product Reviews on Purchasing Decisions. In *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, (pp. 271–278).
- D'Aquino Netto, J. S., Almeida, G. R. T. de, Lobato, F. M. F. & Jacob, Jr., A. F. L. (2019). Melhorando Sistemas de Social CRM por meio de Eletronic Word-of-Mouth. *Revista Eletrônica de Iniciação Científica*, 17(4).
- Einwiller, S. A., & Steilen, S. (2015). 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*, 41(2), 195–204.
- Ernala, S. K., Labetoulle, T., Bane, F., Birnbaum, M. L., Rizvi, Asra F., Kane, J. M. & Choudhury, M. de (2018). Characterizing Audience Engagement and Assessing Its Impact on Social Media Disclosures of Mental Illnesses. In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM)*, (pp. 62–71).
- García, S., Luengo, J., Herrera, F., García, S., Luengo, J., & Herrera, F. (2015). Data Preprocessing in Data Mining. In *Intelligent Systems Reference Library* (Vol. 72).
- Gavilanes, J. M., Flatten, T. C., & Brettel, M. (2018). Content Strategies for Digital Consumer Engagement in Social Networks: Why Advertising Is an Antecedent of Engagement. *Journal of Advertising*, 47(1), 4–23.

- IBGE - Instituto Brasileiro de Geografia e Estatística (2019a). Estimativas da população com referência a 1º de julho de 2019 (xls). Retrieved December 21, 2019, from: <https://agenciadenoticias.ibge.gov.br/agencia-detalhe-de-midia.html?view=mediaibge&catid=2103&id=3098>
- IBGE - Instituto Brasileiro de Geografia e Estatística. (2019). Downloads. Brasil: IBGE.
- IBGE - Instituto Brasileiro de Geografia e Estatística (2020). Produto Interno Bruto - PIB Retrieved February 4, 2020, from: <https://www.ibge.gov.br/explica/pib.php>
- Kim, A. J., & Johnson, K. K. P. (2016). Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook. *Computers in Human Behavior*, 58, 98–108.
- Li, D., Zamani, S., Zhang, J., & Li, P. (2019). Integration of Knowledge Graph Embedding Into Topic Modeling with Hierarchical. In Proceedings of Cld, 940–950.
- Liu, J., Li, C., Ji, Y. G., North, M., & Yang, F. (2017). Like it or not: The Fortune 500's Facebook strategies to generate users' electronic word-of-mouth. *Computers in Human Behavior*, 73, 605–613
- Lobato, F., Pinheiro, M., & Jacob Jr, A. (2017). *Social CRM: Biggest Challenges to Make it Work in the Real World*. Business Information Systems Workshops, 263, 221–232.
- McIlroy, S., Ali, N., Khalid, H., & E. Hassan, A. (2016). Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews. *Empirical Software Engineering*, 21(3), 1067–1106.
- Movel_Pessoal - Files - ownCloud. (n.d.). Retrieved February 5, 2020, from https://cloud.anatel.gov.br/index.php/s/TpaFAwSw7RPfBa8?path=%2FMovel_Pessoal
- Mujahid, S., Sierra, G., Abdalkareem, R., Shihab, E., & Shang, W. (2018). An empirical study of Android Wear user complaints. *Empirical Software Engineering*, 23(6), 3476–3502.
- Nusair, K., Hua, N., Ozturk, A., & Butt, I. (2017). A theoretical framework of electronic word-of-mouth against the backdrop of social networking websites. *Journal of Travel and Tourism Marketing*, 34(5), 653–665.
- Olmedilla, M., Martínez-Torres, M. R., & Toral, S. L. (2016). Harvesting Big Data in social science: A methodological approach for collecting online user-generated content. *Computer Standards and Interfaces*, 46, 79–87.
- Rhee, H. T., & Yang, S. B. (2015). How does hotel attribute importance vary among different travelers? An exploratory case study based on a conjoint analysis. *Electronic Markets*, 25(3), 211–226.
- Rodrigues, L. D. F., Baségio Jr, A. , & Lobato, F. M. F. (2019). Disability-Related News: An Analysis of User-Generated Content on Social Media Posts. In Proceedings of the 16th National Meeting on Artificial and Computational Intelligence.

- Rollins, J. B. (2015). Foundational Methodology for Data Science A 10-stage data science methodology that spans technologies and approaches. *IBM Analytics*.
- Salminen, J., Almerekhi, H., Milenković, M., Jung, S.-G., An, J., Kwak, H., & Jansen, B. J. (2018). Anatomy of Online Hate: Developing a Taxonomy and Machine Learning Models for Identifying and Classifying Hate in Online News Media. In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*, (ICWSM), (pp. 330–339).
- Silva, J. R. da, Brasil, C. C. P., Silva, R. M. da, Brilhante, A. V. M., Carlos, L. M. de B., Bezerra, I. C., & Filho, J. E. de V. (2018). Redes Sociais e Promoção da Saúde: Utilização do Facebook no Contexto da Doação de Sangue. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (30), 107–122.
- Schafer, F., Zeiselmaier, C., Becker, J., & Otten, H. (2019). Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes. In *2018 IEEE International Conference on Technology Management, Operations and Decisions, ICTMOD 2018*, (pp. 190–195).
- Schmäh, M., Wilke, T., & Rossmann, A. (2017). Electronic Word-of-Mouth: A Systematic Literature Analysis. *Lecture Notes in Informatics (LNI)*, 147.
- Sharma, R., Fantin, A. R., Prabhu, N., Guan, C., & Dattakumar, A. (2014). Digital literacy and knowledge societies: A grounded theory investigation of sustainable development. *Telecommunications Policy*, 40(7), 628–643.
- Stone, M. (2015). The evolution of the telecommunications industry — What can we learn from it?. *J Direct Data Digit Mark Pract*, 16, 157–165.
- Tang, C., & Guo, L. (2013). Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication. *Marketing Letters*, 26(1), 67–80.
- Trstenjak, B., Mikac, S., & Donko, D. (2014). KNN with TF-IDF based framework for text categorization. *Procedia Engineering*, 69, 1356–1364.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). 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*, xxxx, 1–17.
- Vu, P. M., Nguyen, T. T., Pham, H. V., & Nguyen, T. T. (2016). Mining user opinions in mobile app reviews: A keyword-based approach. In *Proceedings - 2015 30th IEEE/ACM International Conference on Automated Software Engineering, ASE 2015*, (pp. 749–759).
- Wirth, R. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. In *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, (24959), (pp. 29–39).