

UNIVERSIDADE ESTADUAL DO PIAUÍ – UESPI
CAMPUS PROF. ALEXANDRE ALVES DE OLIVEIRA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

GABRIEL LUCAS GALVÃO DA SILVA

**MINERAÇÃO DE DADOS NO FACEBOOK: USO DE UM MÉTODO ADAPTADO
PARA EXTRAÇÃO E ANÁLISE DE DADOS EM FAN PAGES**

PARNAÍBA
2018

GABRIEL LUCAS GALVÃO DA SILVA

**MINERAÇÃO DE DADOS NO FACEBOOK: USO DE UM MÉTODO ADAPTADO
PARA EXTRAÇÃO E ANÁLISE DE DADOS EM FAN PAGES**

Trabalho de Conclusão de Curso submetido ao Curso de Bacharelado em Ciência da Computação da Universidade Estadual do Piauí, campus Professor Alexandre Alves de Oliveira, Parnaíba, Piauí, como parte dos requisitos legais para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. MSc. Rodrigo Augusto Rocha Souza Baluz

PARNAÍBA
2018

S586m Silva, Gabriel Lucas Galvão da
Mineração de dados no Facebook: uso de um método adaptado
para extração e análise de dados em Fan Pages / Gabriel Lucas
Galvão da Silva. – 2018.
73 f. : il.

Monografia (graduação) – Universidade Estadual do
Piauí – UESPI, Bacharelado em Ciência da Computação,
2018.

“Orientador Prof. MSc. Rodrigo Augusto R. Souza Baluz.”

1. Mineração de Dados. 2. Gestão Competitiva.
3. Análise de Sentimentos. 4. Redes Sociais.
5. Fan Pages do Facebook. I. Título.

CDD: 005

GABRIEL LUCAS GALVÃO DA SILVA

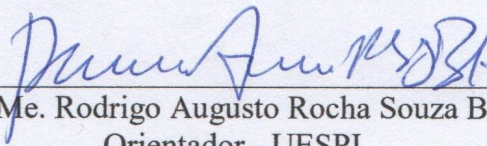
**MINERAÇÃO DE DADOS NO FACEBOOK: USO DE UM MÉTODO ADAPTADO
PARA EXTRAÇÃO E ANÁLISE DE DADOS EM FANPAGES**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Estadual do Piauí – UESPI, Campus Prof. Alexandre Alves de Oliveira, como parte das exigências da disciplina de Estágio Supervisionado, requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

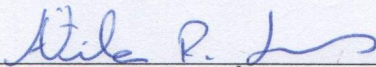
Orientador: Prof. Me. Rodrigo Augusto Rocha Souza Baluz

Monografia Aprovada em: **28 de julho de 2018.**

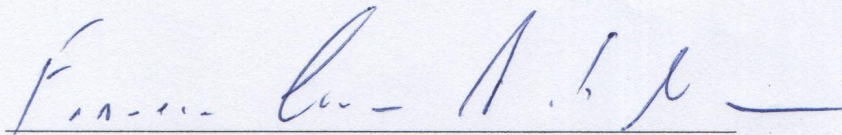
BANCA EXAMINADORA:



Prof. Me. Rodrigo Augusto Rocha Souza Baluz
Orientador - UESPI



Prof. Me. Atila Rabelo Lopes
Avaliador - UESPI



Prof. Me. Francisco Gerson Amorim de Meneses
Avaliador - IFPI

Dedico a meus pais e família, que foram meus incentivadores desde o primeiro ano de faculdade. Aos animais de estimação que me acompanham nos trabalhos caseiros e os meus colegas de classe que estiveram comigo ao decorrer da graduação.

AGRADECIMENTOS

Ao meus pais Cláudia e Expedito por todo o apoio, amor, compreensão, educação e ensinamentos.

A minha tia Amparo pelo amor, carinho e conselhos.

Ao professor Rodrigo Augusto Rocha Souza Baluz, meus sinceros agradecimentos, não apenas pelas orientações firme e segura demonstrada na elaboração deste trabalho, mas também pelo incentivo, confiança e amizade nesses anos de convivência.

A Universidade Estadual do Piauí como um todo, em especial ao Campus Professor Alexandre Alves de Oliveira pela oportunidade de cursar uma graduação e me especializar para o mercado de trabalho.

A minha namorada Débora pelo apoio, carinho e compreensão.

Aos meus amigos Denis, Vinícius, Stéfani e Daniel meu afeto e minha gratidão eterna.

“A sabedoria é um paradoxo. O homem que mais sabe é aquele que mais reconhece a vastidão da sua ignorância”
(Friedrich Nietzsche).

RESUMO

Desde seu surgimento, a Internet tem revolucionado os meios de comunicação servindo de palco para o desenvolvimento de inúmeras aplicações, voltadas para propagação de informações. Entre as plataformas que mais se destacam estão as mídias sociais, como: blogs, fóruns, redes sociais, microblogs entre outras. As redes sociais são um tipo de mídia social, que possibilitam a seus usuários criarem perfis pessoais e interagir com outras pessoas, onde é permitido aos usuários publicarem seus próprios conteúdos ou compartilhar o conteúdo de outras publicações. O *Facebook* é a rede social mais popular do momento, além de permitir às pessoas comuns criarem perfis e interagirem entre si, possibilita que as organizações criem páginas para se relacionar diretamente com o público. Este recurso conhecido com *fan pages*, possibilita as empresas divulgarem seus produtos e serviços de forma gratuita, criar eventos, promover ações publicitárias e muito mais. Todos estes recursos acabam gerando uma série de dados sobre os seguidores das páginas que podem revelar padrões interessantes a serem explorados pelas empresas. Porém devido ao fluxo gigantesco de dados trafegados em uma rede social do porte do *Facebook*, a tarefa de recuperar informações úteis acaba se tornando um tanto quanto complexa. Por este motivo, o presente estudo teve por objetivo criar um método de extração e análise em *fan pages* do *Facebook*, sendo este composto por quatro etapas. A metodologia foi construída baseada nas principais técnicas de extração e análise de dados em redes sociais presentes na literatura e teve como trabalhos norteadores as metodologias propostas por Neves (2013), Hea, Zha e Li (2013), Abrahams et al., (2013), Thomaz (2014) e Santos (2014). Foram utilizadas técnicas de análise de sentimentos em textos, onde foi desenvolvido um script na linguagem de programação python, usando a biblioteca de PLN (Processamento de Linguagem Natural) TextBlod, além da aplicação da técnica de mineração de regras de associação, com a ferramenta *WEKA* e o algoritmo Apriori. Embora o processo tenha explorado o setor de turismo, ele pode ser aplicado em outros tipos de *fan pages*, uma vez que os tipos de dados analisados serão os mesmos. Para verificar o modelo, as técnicas propostas foram aplicadas sobre 3 páginas de agências de turismo no litoral do Piauí, no período de 20 de dezembro de 2017 a 10 de abril de 2018. Foram recuperados 1811 comentários de usuários nas páginas investigadas, onde deste total 42% foram submetidos à análise de sentimentos em textos e 36% foram marcações feitas por outros usuários totalizando 78% de aproveitamento dos dados adquiridos. Dos comentários submetidos a análise de sentimentos 50% foram classificados com neutros, 48% positivos e apenas 2% negativos, com estes resultados pode-se concluir que a imagem do consumidor sobre as três agências analisadas pode ser considerada boa.

PALAVRAS-CHAVE: Mineração de Dados. Gestão Competitiva. Análise de Sentimentos. Redes Sociais. *Fan pages* do *Facebook*.

ABSTRACT

Since its inception, the Internet has revolutionized the media as the stage for the development of numerous applications, aimed at information dissemination. Among the platforms that stand out most are social media, such as: blogs, forums, social networks, microblogs among others. Social media is a type of social media that enables users to create personal profiles and interact with others, where users are allowed to publish their own content or share content from other publications. Facebook is the most popular social network of the day, as well as allowing ordinary people to create profiles and interact with each other, enabling organizations to create pages to relate directly to the public. This resource known as fan pages enables companies to advertise their products and services for free, create events, promote advertising and more. All of these features end up generating a lot of data about page followers that may reveal interesting patterns to be explored by companies. But because of the gigantic flow of data trafficked into a Facebook-sized social network, the task of retrieving useful information turns out to be somewhat complex. For this reason, the present study aimed to create a method of extraction and analysis in Facebook fan pages, which is composed of four stages. The methodology was based on the main techniques of extraction and data analysis in social networks present in the literature and had as guiding works the methodologies proposed by Neves (2013), Hea, Zha and Li (2013), Abrahams et al., (2013), Thomaz (2014) and Santos (2014). We used feelings analysis techniques in texts, where a script was developed in the python programming language, using the textblob PLN (Natural Language Processing) library, as well as the application of the association rules mining technique, with the WEKA tool and the apriori algorithm. Although the process has explored the tourism industry, it can be applied to other types of fan pages, since the types of data analyzed will be the same. To verify the model, the proposed techniques were applied on 3 pages of tourism agencies in the coast of Piau , from December 20, 2017 to April 10, 2018. 1811 comments were retrieved from users in the pages investigated, where of this total 42% were submitted to the analysis of feelings in texts and 36% were markings made by other users totaling 78% of use of the acquired data. Of the comments submitted to the analysis of feelings, 50% were classified as neutral, 48% positive and only 2% negative, with these results it can be concluded that the consumer image about the three analyzed agencies can be considered good.

KEY-WORDS: Data Mining. Competitive Management. Analysis of Feelings. Social Networks. *Fan pages from Facebook.*

LISTA DE FIGURAS

Figura 1. Abordagens de descoberta de conhecimento.	15
Figura 2. Processo de descoberta de conhecimento em base de dados.	16
Figura 3. Redes sociais mais populares ao redor do mundo.....	24
Figura 4. Número de usuários ativos mensais no <i>Facebook</i>	26
Figura 5. Metodologia de coleta e análise de tweets proposta por Neves.	31
Figura 6. Metodologia de mineração conteúdo texto em mídias sociais.....	32
Figura 7. Processo de isolamento de componentes de veículos em mídias sociais.....	34
Figura 8. Processo de mineração de conteúdo em mídias sociais para auxiliar na gestão de destinos turísticos.	35
Figura 9. Metodologia da pesquisa.....	38
Figura 10. Metodologia proposta para extração e análise de dados em <i>fan pages</i> do <i>Facebook</i>	42
Figura 11. Tela inicial da ferramenta de coleta NETVIZZ.	48
Figura 12. Tela de busca por dados de página no NETVIZZ.....	49
Figura 13. Saída de dados após a descompactação.	49
Figura 14. Script em python para análise de polaridade e subjetividade em postagens de comentários no <i>Facebook</i>	52
Figura 15. Execução do algoritmo a priori na ferramenta <i>Weka</i>	53
Figura 16. Emoticons presentes no <i>Facebook</i>	54
Figura 17. Visualização de uma rede antes de ser tratada na ferramenta gephi.	55
Figura 18. Porcentagem de cada <i>fan pages</i> do total de postagens recuperadas.....	56
Figura 19. Percentual de comentários analisados.	57
Figura 20. Sentimento percebido nos comentários pela análise de sentimentos em textos.....	59
Figura 21. Análise de sentimentos das agências presentes no litoral do piauiense.	60
Figura 22. Nuvem de palavras dos comentários da Clip Turismo.....	63
Figura 23. Nuvem de palavra mais frequentes nos comentários dos usuários na página da CVC Parnaíba.	64
Figura 24. Nuvem de palavras geradas pelos comentários na <i>fan pages</i> da Eco Adventure....	64
Figura 25. Rede de relacionamento da Clip Ecoturismo.	65
Figura 26. Rede de relacionamento da Eco Adventure.	66

LISTA DE QUADROS

Quadro 1. Categorias e características das mídias sociais.....	21
Quadro 2. Conceitos importantes sobre extração de conhecimento em base de dados.....	29
Quadro 3. Características da pesquisa quanto ao propósito natureza dos dados delineamento e natureza da pesquisa.	37
Quadro 4. Contribuição dos autores para elaboração do método de extração e análise em <i>fan pages</i> do <i>Facebook</i>	40
Quadro 5. Etapas e processos da metodologia de extração e análise de <i>fan pages</i>	41
Quadro 6. Maiores agências de turismo em número de seguidores no litoral piauiense.....	47
Quadro 7. Etapa de limpeza dos comentários.....	50
Quadro 8. Atributos, significados e valores para a leitura das regras de associação geradas. .	53
Quadro 9. Conversão dos dados numéricos em nominais.	54
Quadro 10. Números gerais de comentários analisados, marcações e postagens inválidas.	58
Quadro 11. Valores retornados pelo script de análise de sentimentos em textos.....	59
Quadro 12. Top 10 regras de associação na página da Clip Ecoturismo.....	61

SUMÁRIO

1 INTRODUÇÃO	12
2 REVISÃO DA LITERATURA	15
2.1 DESCOBERTA E GESTÃO DO CONHECIMENTO	15
2.2 MÍDIAS SOCIAIS: HISTÓRICO, CONCEITOS E APLICAÇÕES.....	18
2.3 FACEBOOK COMO FERRAMENTA PARA GESTÃO COMPETITIVA	25
2.4 MINERAÇÃO DE DADOS EM REDES SOCIAIS.....	28
2.4.1 Modelo proposto por Neves (2013).....	30
2.4.2 Modelo proposto por Hea, Zha e Li (2013).....	32
2.4.3 Modelo proposto por Abrahams et al., (2013).....	33
2.4.4 Modelo proposto por Thomaz (2014).....	34
2.4.5 Modelo proposto por Santos (2014).....	36
3 MATERIAIS E MÉTODOS DA PESQUISA	37
3.1 CARACTERIZAÇÃO DA PESQUISA	37
3.2 AMBIENTE DE PESQUISA.....	37
3.3 PROCEDIMENTOS METODOLÓGICOS.....	38
3.4 CONSTRUÇÃO DO MODELO DE EXTRAÇÃO E ANÁLISE DE DADOS EM FAN PAGES	39
3.5 FERRAMENTAS	42
3.5.1 Python E Textblod.....	43
3.5.2 Netvizz.....	43
3.5.3 Tagul.....	44
3.5.4 WEKA	44
3.5.5 Gephi	45
4 PROPOSTA DE APLICAÇÃO: FACEBOOK NA GESTÃO DO TURISMO NA REGIÃO DO LITORAL PIAUIENSE.....	46
4.1 ETAPA 1 – COLETA DE DADOS	47
4.2 ETAPA 2 - PRÉ-PROCESSAMENTO.....	50
4.3 ETAPA 3 – PROCESSAMENTO	51
4.4 ETAPA 4 – ANÁLISE E VISUALIZAÇÃO OS RESULTADOS.....	54
5 ANÁLISE E DISCUSSÃO DOS RESULTADOS	56

5.1 COLETA DE DADOS	56
5.2 RESULTADOS DOS PROCESSOS DE MINERAÇÃO	58
5.3 VISUALIZAÇÃO DA REDE DE RELACIONAMENTO.....	62
6 CONSIDERAÇÕES FINAIS.....	67
REFERÊNCIAS	68

1 INTRODUÇÃO

A contemporaneidade é brindada por uma era de avanços tecnológicos em uma velocidade vertiginosa. Tal característica social atribui aos tempos atuais, o paradigma de sociedade da informação, uma cultura globalizada em que a informação ocupa o topo da lista dos bens mais preciosos produzidos pela humanidade em nosso tempo. Esta característica de sociedade possibilita uma série de oportunidade, mais também inúmeros desafios para profissionais e organizações de buscam se destaca em um mercado globalizado cada vez mais competitivo.

Para Gama et al., (2016) este contexto vivenciado pela sociedade moderna, é caracterizado pela gigantesca quantidade de informações que estão acessíveis a todo momento, esta realidade acaba por demandar habilidades dos profissionais e das empresas que vão muito além do que apenas deter as informações, tornando-se essencial saber a melhor maneira de aproveitá-las para geração de conhecimento útil para a maximização dos resultados positivos dentro das organizações.

Nestes cenários de avanços tecnológicos constantes, a Internet tem atuado como fonte primordial de informações, abrigando uma série de aplicações que tem revolucionado a maneira como as pessoas e organizações se relacionam. Entre essas aplicações que tem ganhado bastante destaque no âmbito acadêmico e corporativo estão as mídias sociais. As redes sociais são uma das categorias de mídias sociais mais utilizadas na atualidade, um exemplo de redes social bastante conhecida e explorada é a rede social *Facebook*, que na atualidade lidera o ranking das redes sociais mais acessadas do mercado (NEVES; MACHOTA, 2016).

Entretanto, as informações mais relevantes para geração de conhecimento útil nem sempre estão expostas explicitamente nas redes sociais. Para utilizasse todo o potencial informacional que as redes sociais possuem, é preciso minerar estas informações para a descoberta efetiva e funcional deste conhecimento. Percebendo este contexto de mudanças de paradigmas, que as organizações têm aderido cada vez mais as redes sociais, não somente como ferramenta para o marketing digital, mais para entender o comportamento do consumidor e construir um relacionamento duradouro e de qualidade.

O *Facebook* além de ser a rede social com o maior número de usuários ativos na atualidade, também possibilita às organizações a interação com seus clientes por meios das suas *fan pages* (Páginas de fãs). Este recurso da rede social permite que as organizações criem um grupo de pessoas que engajadas com a sua empresa, consolidando a imagem e o conceito da marca perante o mercado (AGUIAR, 2016).

As *fan pages* possibilitam que as organizações divulguem seus produtos e serviços, atraírem novos clientes e verificarem nichos de mercados, gerando aumento da visibilidade e presença da marca na vida do consumidor (REZ, 2016). Estas funcionalidades das *fan pages* acabam gerando grandes quantidades de dados sobre os seguidores das páginas, muitos desses dados podem ser úteis para as organizações direcionarem melhor as suas estratégias de marketing, além de desenvolver produtos melhores que satisfaçam os desejos dos clientes. Porém é preciso encontrar formas de identificar e mensurar os dados que podem gerar conhecimento sobre os usuários para as organizações no ambiente das *fan pages*.

Observando este cenário, fica evidente que adquirir e transformar uma gigantesca quantidade de dados que tem como fonte uma rede social como o *Facebook* representa um problema para as organizações. Tendo isso em mente o corpo do presente estudo busca responder: como desenvolver um método capaz de extrair e analisar dados das *fan pages* do *Facebook*.

Uma alternativa para extrair o conhecimento das *fan pages* e através da utilização de técnicas de descoberta de conhecimento, como mineração de dados e análise de sentimentos em textos. Os dados extraídos das páginas, estão relacionados às atividades dos usuários que passaram pelas páginas e deixaram informações importantes que contêm padrões relevantes de serem analisados no contexto completo da página. A utilização da análise de sentimentos em textos, vem ajudando a compreender as emoções dos usuários a respeito de algum determinado assunto tornando se bastante relevante, quanto se pensar em distinguir os comentários positivos dos negativos nas páginas das organizações.

Outra técnica de grande valia para mensurar as opiniões presentes em comentários dos usuários, é a análise visual de redes sociais por meio de nuvens de palavras. Esta técnica permite destacar os termos mais utilizados pelos usuários nos comentários além de revelar outro não tão evidentes mais que aparecem secundariamente entre os principais, revelando tendências, desejos e opiniões.

Como objetivo geral deste estudo foi criar uma metodologia que reunisse diversas técnicas para realizar as tarefas de extração e análise de dados em *fan pages* da rede social *Facebook* objetivando auxiliar as organizações no processo de gestão e tomada de decisões estratégicas. Visando alcançar o objetivo geral, foram propostos alguns objetivos específicos, como: (a) identificar modelos e ferramentas de mineração de dados em redes sociais presentes na literatura; (b) criar um modelo para coleta e análise de dados em *fan pages* do *Facebook*; (c) validar o modelo proposto por meio da aplicação real sobre dados gerados por usuários da rede social *Facebook*.

Para a elaboração da metodologia proposta, foram estudadas as principais técnicas de análise de redes sociais presentes literatura e identificadas as que mais se adequa para a utilização nas *fan pages*. Como os dados textuais foram os mais presentes nas páginas, foi elaborado um script para analisar os sentimentos dos comentários dos usuários feitos nas páginas em relação postagens. Os dados estáticos com número de curtidas, comentários, reações entre outros foram transformados em nominais para geração de regras de associação para descobrir em quais padrões são mais vantajosos em termos de postagens pelas páginas para alcançar inúmeras visualizações.

Além deste capítulo de introdução, que buscou expor a temática deste estudo, a estrutura deste documento foi composta da seguinte maneira:

No Capítulo 2, “Revisão de Literatura”, serão apresentados os conceitos para a compreensão da temática explorada pelo trabalho onde será feita a revisão sobre descoberta e gestão do conhecimento, mídias sociais, *Facebook* como ferramenta para gestão competitiva e os modelos de mineração de dados em redes sociais.

No Capítulo 3, “Materiais e Métodos da Pesquisa”, serão apresentados as características da pesquisa, o ambiente no qual se deu o desenvolvimento, a construção do modelo para extração e análise de dados em *fan pages* e as ferramentas utilizadas.

O Capítulo 4, “Proposta de Aplicação: *Facebook* na Gestão do Turismo na Região do Litoral Piauiense”, serão apresentados o ambiente escolhido para aplicar a metodologia desenvolvida, apresentar os procedimentos e as configurações das ferramentas utilizadas em cada uma das etapas.

No Capítulo 5, “Análise e Discussão dos Resultados”, serão apresentados os principais resultados da análise dos sentimentos nos comentários em cada *fan pages*, bem como as principais regras geradas, as nuvens de palavras e a rede de relacionamentos de cada página.

O Capítulo 6, “Considerações Finais”, apresenta as conclusões e os trabalhos futuros sugeridos.

2 REVISÃO DA LITERATURA

2.1 DESCOBERTA E GESTÃO DO CONHECIMENTO

Com a popularização da internet e das mídias sociais, o volume de dados trafegados pela rede aumenta na mesma proporção que mais usuários as utilizam. Este aumento na quantidade de tráfego de dados gerou a necessidade de transformar dados brutos em conhecimento, uma vez que muitos dados podem se mostrar úteis, mas se apresentam de maneira implícita em meio a uma vastidão de dados. Alinhado com este pensamento Longhi (2016), expressa que é possível gerar conhecimento e assistir o processo de tomada de decisão em organizações a partir desses dados presentes na internet desde que eles estejam bem estruturados.

A descoberta de conhecimento ou *Knowledge Discovery* (KD), surge como uma técnica de análise de dados, que buscam auxiliar na aquisição de novos conhecimentos relevantes que não foram apresentados anteriormente e possivelmente ocultos baseado na manipulação dados massivos (MORAIS; AMBRÓSIO, 2007). Para Morais e Ambrósio (2007), a descoberta de conhecimento pode se apresentar de dois modos distintos, que variam entre estruturados e não estruturados conforme mostra a Figura 1.

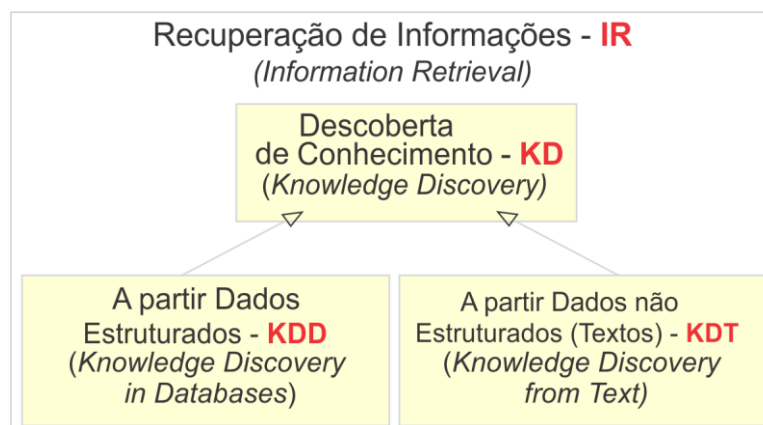


Figura 1. Abordagens de descoberta de conhecimento.
Fonte: (Morais e Ambrósio (2007)).

O primeiro modo de KD, é referente ao processo de extração de conhecimento em base de dados, do inglês *knowledge-discovery in databases* (KDD), consiste na aplicação de algoritmos específicos para mineração em bases de dados, como o propósito de revelar informações, associações relevantes e padrões extraídos. Este modelo de descoberta conhecimento surgiu buscando solucionar o problema da sobrecarga de dados gigantesca

quantidade de dados trafegados na rede pela assim chamada “era da informação” (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

O escopo básico do processo de KDD, se constitui de várias etapas, dentre estas os processos que envolvem a preparação dos dados, limpeza, extração de padrões, avaliação e análise dos resultados recuperados para determinar quais padrões são realmente úteis e podem ser considerados novos conhecimentos. Estes processos são primordiais para garantir que os resultados gerem conhecimentos úteis ao término (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; SANTOS, 2009).

O processo de KDD apresentado por Fayyad, Piatetsky-Shapiro E Smyth (1996), é ilustrado na Figura 2. Composto-se de cinco etapas, sendo estas: Seleção dos Dados; Pré-processamento e Limpeza dos Dados; Transformação dos Dados; Mineração de Dados; Interpretação e Avaliação dos Resultados.

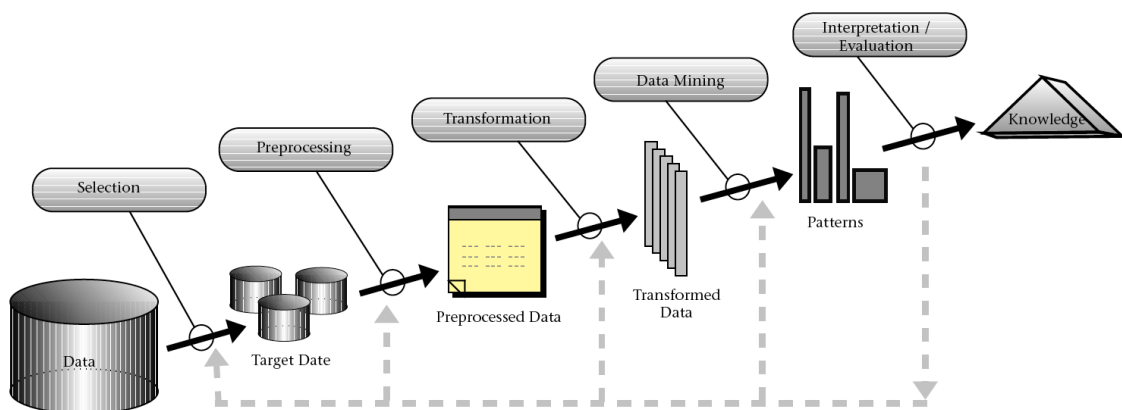


Figura 2. Processo de descoberta de conhecimento em base de dados.
Fonte: Fayyad, Piatetsky-Shapiro e Smyth (1996).

Conforme a Figura 2, as etapas do processo de descoberta de conhecimento podem ser descritas como:

- **Seleção:** consiste em identificar os dados, subconjuntos de variáveis e amostragem de dados das quais pretende extrair informações.
- **Pré-processamento:** é o processo no qual se busca limpar os dados, executando tarefas para remoção de ruídos, além de identificar e remover os valores inconsistentes.
- **Transformação:** é realizado a transformação dos dados através de redução de dimensionalidade, levando em consideração os objetivos definidos no início do processo.

- **Mineração de dados:** consiste na aplicação de técnicas computacionais e algoritmos de mineração de dados visando encontrar padrões e modelos.
- **Avaliação:** etapa de pós-processamento, consiste na interpretação e avaliação dos resultados atingidos, pode representar a possibilidade de retornar a uma etapa anterior do processo.

O segundo modo de KD, consiste no processo de descoberta de conhecimento em texto segundo. Para Silva (2012), a técnica de *Knowledge Discovery in Text* (KDT) engloba fases semelhantes ao KDD em especial a parte que diz respeito a mineração e avaliação. Porém existem algumas divergências, mesmo que possuam processos idênticos, sua aplicabilidade difere, uma vez que o KDT utiliza-se de técnicas específicas para bases de textos visando obter o melhor tratamento possível para gerar inteligência sobre os dados textuais (BARION, 2008).

Segundo Pal, Talwar e Mitra (2002) o fato de que os dados textuais não dispõem de uma estrutura bem definida como as de um registro de banco de dados por exemplo, a tarefa de mineração exige processos adicionais, como criar uma estrutura com os dados e reestruturar os elementos para uma forma que seja legível para as máquinas mediante a aplicação de técnicas como marcações (*tags*). Utilizando a mineração de texto é possível estruturar os dados textuais e sistematizar a tarefa de identificar, extrair, integrar e explorar novos conhecimentos de forma instintiva (TSURUOKA; TSUJII; ANANIADOU, 2008)

Devido às características das técnicas empregadas na mineração de texto, a consulta a especialistas não se faz necessária, pois de modo geral são utilizados algoritmos que aprendem de maneira não supervisionada a coletar as informações relevantes de um texto. O emprego dessas técnicas de agrupamento de documentos vem facilitando toda a cadeia de processos envolvidos na mineração de textos (CORRÊA, 2012).

Para a gestão do conhecimento, o emprego das técnicas de KD vem para agregar valor dentro das organizações uma vez que, é nitidamente perceptível para a sociedade que seu desenvolvimento depende intrinsecamente das informações e do conhecimento gerados em seu âmbito, e engloba todos os setores nela inseridos. A partir do século XX, esta premissa se tornou mais evidente e contribuiu para a evolução de todas as áreas que trabalham com o conhecimento e em sua geração. Tais fatores estão diretamente ligados à economia nos mais diversos níveis, e podem ser percebidos quando olhamos para as organizações e conseqüentemente para sua gestão (JANNUZZI; FALSARELLA; SUGAHARA, 2016).

Neste cenário a importância de gerenciar o conhecimento torna-se valiosíssima e fundamental para a tomada de decisões estratégicas dentro do ambiente corporativo e na vida

dos gestores das organizações. Ao longo da história a boa gestão do conhecimento vem atuando como papel essencial, onde deter o conhecimento e aplicá-lo efetivamente impulsionaram diversas sociedades a conquistas diversas outras, simplesmente por entender que gerir as informações e utilizá-las de forma estratégicas representavam e representam vantagens competitivas (LUCHESE, 2012).

Este conceito tornou-se bem mais evidente com a globalização, a cada novo avanço da sociedade o mercado global fica cada vez mais competitivo e exigente por excelência. Isto tem influências diretas nas organizações, definindo comportamentos, segmentos e na busca incessante por inovações que possam agregar valor efetivo dentro das empresas. Todo este processo de modificação nos modelos de gestão dentro das organizações, acabou por causa a ruptura de diversos paradigmas e conseqüentemente acaba por criar novos, tudo isso impulsionado pela era do conhecimento (LONGO et al., 2014).

Porém, é preciso ressaltar que para a eficácia do processo de gestão do conhecimento ocorra trazendo benefícios efetivos para as organizações, é preciso primeiramente criar um ambiente que seja favorável, por meio de treinamentos, adquirindo ou desenvolvendo talentos dentro da equipe, estabelecendo parcerias benéficas e desenvolvendo pesquisas que colaborem com o aperfeiçoamento e aquisição de conhecimento dentro da organização (DUQUE; COSTA, 2016).

O fato é que em tempos de globalização a gestão do conhecimento se tornou não apenas um diferencial competitivo dentro das organizações mais sim uma questão de sobrevivência no mercado cada vez mais competitivo. Para Longo et al., (2014) o papel do conhecimento para as organizações hoje, tornou-se mais importante do que os processos produtivos. Nesse contexto Duque e Costa (2016) expressam que nesse novo paradigma da sociedade, às empresas são forçadas a desenvolver-se tecnologicamente não somente nos aspectos de acesso a informação, mais na capacidade de manusear as informações adquiridas e transformá-las em conhecimento.

2.2 MÍDIAS SOCIAIS: HISTÓRICO, CONCEITOS E APLICAÇÕES

O surgimento das redes sociais precede o da tecnologia na era digital, se estabelecendo como atividade humana (SAFKO; BRAKE, 2010). Porém, adota-se perante a postura teórica do presente estudo, seguindo o pensamento de Torres (2009) e Safko e Brake (2010), que assumiam as mídias sociais como um conglomerado de diferentes tipos de mídia, que por sua

vez engloba o conceito de redes sociais, onde na verdade as redes sociais se apresentam com uma categoria no universo das mídias sociais.

Diversos autores buscam definir um conceito para as mídias sociais, avaliando diferentes métricas afins de estabelecer uma definição clara e consistente (O'REILLY, 2005; SONG, 2010). Para (SANTOS et al., 2015), às mídias sociais são aplicações construídas a partir da tecnologia da *Web 2.0* voltadas para internet, que visam permitir a criação e troca de conteúdo entre os usuários, e que como um todo se constituem como um conjunto de diferentes tipos de meios de comunicação, onde podem ser incluídos meios tradicionais como jornal, rádio, televisão, revistas e meios não tradicionais redes sociais, blogs, fóruns de discussão e etc.

Na visão de Torres (2009), as mídias sociais são website que proporciona aos usuários atuarem como produtores e ou consumidores de conteúdo, criando e compartilhando informações de e para pessoas. Isto se caracteriza pela própria natureza da *Web* como ambiente livre e democrático, que possibilita a socialização entre pessoas em comunidade, em um ambiente virtual, com a finalidade de compartilhamento de conteúdo em forma de mídias digitais, admitindo a colaboração e atuação dos usuários e suas práticas sociais.

Complementar a esta ideia Safko e Brake (2010) ressaltam os aspectos que tangem as atividades, práticas e comportamentos realizados por grupos de pessoas reunidas online, com objetivo de compartilhar informações, conhecimentos e opiniões através de aplicações web que permitam a conversação entre os usuários e transmitam o conteúdo pelos usuários em forma de palavras, imagens, vídeos e áudios. Abdel-Hafez e Xu (2013) acrescentam que os processos realizados nas plataformas de mídia social devem se interativos e facilitar a comunicação entre os usuários.

Em seu trabalho Abrahams et al., (2013), discorre sobre os serviços online que as mídias sociais oferecem aos usuários, como fóruns públicos e listas de discussão, *wikis*, redes sociais, comunidades e grupos públicos, análises de produtos dos clientes, comentários públicos, artigos e notícias gerados pelos usuários, micro blogs dentre outros que fazem parte do universo das redes sociais, ressaltando formas descentralizadas de criação, edição e publicação de conteúdo pelos usuários.

Para Kietzmann et al., (2011), as mídias sócias consistem no emprego das tecnologias móveis e baseadas na Internet para proporcionar sistemas interativos que possibilitem aos indivíduos, comunidades e organizações interagir, cocriar, discutir, publicar e modificar conteúdo. De fato, as mídias sociais se tornaram não somente um meio de interação entre usuários isolados mais, possibilitam a organização de grupos com interesses em comum e as organizações se comunicarem com os consumidores em plataformas digitais.

Analisando a percepção dos autores, pode-se abstrair algumas características importantes acerca das mídias sociais, das quais envolvem a utilização de plataformas e aplicações na Internet, facilidade de interação e comunicação entre usuários, grupos e comunidades de usuários e organizações privadas, utilização de dispositivos móveis como computadores, notebooks, tablets e smartphones, atividades que envolvem a comunicação, relacionamento, criação, colaboração, discussão, produção de conteúdo, entretenimento, expressar opinião e conteúdo em diversos formatos de mídia como textos, imagens, vídeos e áudios.

Segundo Thomaz (2014) destaca o fato de que existem diversas mídias sociais para um vasto número de categorias, das quais os formatos de mídia variam de acordo com o tipo de aplicação. Para Kaplan e Haenlein (2010), ainda falta um modelo sistemático que possa classificar as mídias sociais, pois novas aplicações surgem constantemente com os avanços da tecnologia e sugeriram um padrão de classificação sistêmica constituída por seis categorias: colaboração, blogs, comunidades de conteúdo, redes sociais, jogos e mundos virtuais.

(SAFKO, 2012) categorizou e classificou as mídias sociais e seus estudos, apresentados no Quadro 1.

CATEGORIA	CARACTERIZAÇÃO
Redes Sociais	<i>LinkedIn, MySpace e Facebook</i> são exemplos desta categoria de mídias sociais onde a participação é pré-requisito para aumentar a rede de contatos.
Publicação	<i>Wordpress, Wikipedia ou Blogger</i> , são os principais nomes desta categoria onde se privilegia a publicação de conteúdo, independentemente da sua finalidade ou origem.
Fotografias	Espaços de publicação de fotos online como o <i>Flickr</i> ou o <i>Picasa</i> representam este tipo de mídia social.
Áudio	Os <i>podcasts</i> e o <i>iTunes</i> assumem-se como os principais componentes desta categoria onde a disseminação de conteúdo através do som é o que a caracteriza
Vídeo	Produzir e publicar conteúdos aliando imagem e som, são ações que instrumentos como o <i>YouTube, Vimeo</i> ou <i>Metacafe</i> permitem concretizar por meio desta mídia social.
<i>Microblogging</i>	Representa a categoria onde se criam e disseminam mensagens de texto curtas aliadas de alguma interatividade. O <i>Twitter</i> e o <i>Tumblr</i> são dois dos seus principais representantes.
<i>Livecasting</i>	Permitem a transmissão de conteúdo ao vivo pela internet de qualquer lugar <i>Justin.tv</i> e <i>TalkShoe</i> são exemplos de mídias de <i>livecasting</i> .
Mundos Virtuais	Exemplificados por mídias tais como o <i>Second Life</i> onde os seus participantes assumem identidades, papéis e atividades, semelhantes ou totalmente diferentes daquelas que possuem na vida real.

Jogos	Por meio de mídias como o <i>Farmville</i> , usuários de qualquer lugar podem interagir com outros usuários por meio da participação em jogos online.
RSS	Primeiramente conhecido como <i>Rich Site Summary</i> , o atual <i>Really Simple Syndication</i> permite ao usuário ser notificado sobre conteúdos de sua escolha no momento que eles são publicados, sem precisar acessar diretamente os locais que o disponibilizam.
Agregadores	Permitem reunir, atualizar e armazenar conteúdos de forma a que sejam facilmente acessíveis a qualquer momento. <i>TweetDeck</i> e <i>Hootsuite</i> são exemplos desta categoria
Produtividade	Representam uma solução em que o usuário, por meio de um único clique, consegue encaminhar todo o seu conteúdo, no momento em que o publica, para todos os seus seguidores. Exemplo: <i>Feedreader</i> ou <i>iGoogle</i>
Busca	Mídias como o Google que permitem ao usuário procurar conteúdo de seu interesse a qualquer momento e que esteja disponível online
Interpessoal	Aqui estão inseridas ferramentas como o <i>Skype</i> ou o <i>MSN Messenger</i> que facilitam a comunicação direta entre usuários, melhorando inclusive a colaboração entre si.

Quadro 1. Categorias e características das mídias sociais.

Fonte: Adaptado de Safko (2010, *apud* NEVES, 2013, p.40).

Ao observar as categorias e respectivamente suas características apresentadas no Quadro 1, e as observadas Kaplan e Haenlein (2010), pode-se concluir que a evolução e o surgimento de novas definições de mídia social ocorrem de maneira muito rápida, impulsionadas pelos avanços nas tecnologias de comunicação e informação. Estas mudanças possibilitam novas oportunidades e o surgimento de inúmeras aplicações afetando desde a forma com que um usuário comum se comunica até as grandes corporações que se utilizam das mídias sociais com as mais variadas finalidades.

Entre as várias aplicações que se manifestaram a partir destas mudanças, destaca-se as redes sociais, porém para compreendermos o seu conceito, precisamos primeiro abstrair o que uma “rede” (*network*), embora o conceito não se limita apenas aos que serão apresentados nesse trabalho, na verdade ele vem adquirindo vários significados ao longo dos anos, mas podemos atribuir quanto ao propósito da presente pesquisa como sendo “rede”: um conjunto de nodos e seus elos; um conjunto de objetos ou pessoas interligadas uns aos outros; um sistema que permita transitar elementos materiais ou imateriais entre as entidades; uma comunidade que não esteja geograficamente próxima. Partindo desses conceitos podemos conceber que a definição de rede social descende destas ideias, podem ser definidas como um conjunto de usuários que possuem autonomia para compartilhar seus valores e interesses, podendo ou não possui ideias em comum (MARTELETO, 2001).

Podemos complementar que o conceito de rede social pode ser empregado de forma diferente a cada caso, variando de acordo com os dados e o relacionamento entre os indivíduos ou elementos envolvidos. Embora as redes sociais sejam estudadas a vários anos, concorda-se que não existe uma teoria sobre elas, mais podemos empregar sobre as mesmas diversas teorias sociais, que necessitam de clareza nos dados onde seus elementos e relacionamentos estejam bem definidos.

Para Emirbayer e Goodwin (1994), uma rede social pode ser definida como um grafo, orientado ou não, que representa uma realidade ou um micromundo, no qual os vértices descrevem os indivíduos ou classes de indivíduos e as arestas representam os relacionamentos existentes entre esses indivíduos. Um ou mais atributos podem compartilhar os mesmos relacionamentos, apresentando uma característica multirelacional, dos quais os relacionamentos podem ser unidirecionais e não necessariamente precisam ser binários, e as sociais podem ser compreendidas como fontes de dados heterogêneos.

Existem algumas formas de rotular uma rede social, das quais são normalmente classificadas em três categorias: redes aleatórias, redes de mundo pequeno e redes sem escala. Para redes aleatórias, existem dois modelos rigorosamente associado com a criação de grafos aleatórios. O primeiro foi introduzido por Erdős e Rényi (1960), criadores do modelo de redes aleatórias, onde os nodos são conectados de maneira aleatória possuindo aproximadamente a mesma quantidade de conexões. Para gerar o modelo deve-se detalhar o número de nós no grafo gerado e qual a probabilidade de uma aresta entre dois nós ser formada, onde as arestas não dependem uma das outras. O segundo modelo, trouxe à tona o conceito de redes igualitárias, isso pelo fato de que os nodos possuem mais ou menos o mesmo número de ligações (BARABÁSI, 2003).

A redes de pequeno mundo consiste em um grafo matemático onde grande parte das ligações é estabelecida entre os vértices mais próximos, na qual a distância média entre dois vértices não ultrapassa um número pequeno de vértices (WATTS; STROGATZ, 1998). O conceito de redes pequeno mundo foi trazido com o experimento de Milgram (1963), cujo objetivo foi verificar o grau de separação entre as entidades (pessoas) da rede social proposta. O experimento consistia no envio de cartas aleatoriamente a várias pessoas, pedindo-lhes que as enviasse a um alvo específico onde, caso não o conhecessem, estas deveriam ser enviadas a outra pessoa, a fim de ser repassada pelos nodos da rede até que o alvo fosse atingido. Ao final do experimento, das cartas que chegaram ao alvo final, passaram por um pequeno número de pessoas. Esse estudo ficou conhecido como “fenômeno do mundo pequeno” (*Small-World*

Phenomenon), ou popularmente como “princípio dos seis graus de separação” (DEGENNE; FORSÉ, 1999).

Esse paradigma pode facilmente ser observado nas redes sociais, onde cada indivíduo tem um grupo de pessoas com as quais se relacionam diretamente ao redor do mundo, que também possuem seu grupo, fazendo com que as pessoas estejam conectadas entre si, apenas com um pequeno grau de separação. Foram elaborados outros conceitos ao decorrer do estudo, Granovetter (1983) criou os conceitos de laços fracos (*Weak ties*), e laços fortes (*Strong ties*), onde os laços fracos são considerados dinâmicos, aqueles que se desfazem mais facilmente, já os laços fortes são laços estáticos, por serem mais constantes. Durante seu estudo Granovetter (1983) concluiu que os laços fracos são mais importantes para a sustentação de uma rede social, por conectarem indivíduos de diversos grupos sociais fazendo com que essa estrutura faça parte de uma rede social de fato (CERVI, 2008).

Milgram assim como Erdős e Rényi referiam-se às redes sociais como redes aleatórias, ou seja, em seu entendimento assumiram que as conexões entre os nós da rede se formavam de maneira randômica, já para Barabási (2003) as redes não se formam de maneira aleatória, porém existe uma ordem dinâmica na estruturação e na forma como a rede se organiza. Este padrão foi observado por Barabási (2003), que apontou para o fato de que quanto maior for número de conexões de um nó, mais chances ele terá de fazer novas conexões no futuro, indicando a existência de nós altamente conectados (chamados de hubs ou conectores), enquanto outras possuem poucas conexões. Ele chamou essa característica de "*preferential attachment*" (conexão preferencial), no qual um novo nó inserido na rede tende a conectar-se aos nós pré-existentes que fossem mais densos.

Para Benevenuto, Almeida e Silva, (2011), uma rede social pode ser definida como um serviço Web que possibilita aos usuários: criar perfis públicos ou semipúblicos dentro de um sistema; unir-se a uma lista de outros usuários das quais o usuário compartilha conexões e; visualizar e percorrer listas de conexões, suas ou de outros usuários do sistema. Uma rede social online também poder ser entendida como um grafo, incorporando características importantes, e usando-se deste modelo é possível prever como a uma rede se comporta no futuro.

Outra característica das redes sociais importe a ser ressaltada e a privacidade, a circulação de dados pessoais, familiares, hábitos, preferências entre outra infinidade de informações que se pode compartilhar no ambiente online podem abrir espaço para situações perigosas para os usuários. Existem diversas campanhas e materiais de conscientização do uso moderado da internet, e apesar dos estudos avançarem nessa área, não é possível avaliar os riscos versus os benefícios para os usuários (OLIVEIRA, 2012).

Existem diversas redes sociais online disponíveis na Web, onde suas particularidades variam de acordo com seus principais objetivos, como por exemplo rede de amizade como Facebook, compartilhamento de vídeos como YouTube, troca de mensagens como *WhatsApp* e *Twitter*, redes de contatos profissionais com *LinkedIn* entre outros tipos de redes.

As redes sociais mais populares geralmente apresentam suporte a vários idiomas, isso permite que os usuários possam se conectar a várias pessoas ou amigos rompendo a fronteira geográfica, política ou econômica. Cerca de 2 bilhões de usuários que estão conectados à internet possuem conta em alguma rede social, sendo que a expectativa para os próximos anos é que esse número aumente ainda mais com a popularização dos dispositivos e redes móveis (STATISTICS, 2018a).

Seguindo a ideia de Marteleto (2001), que estudar a natureza das redes sociais pode ser justificado pela evidência de dados que compõem a realidade da sociedade contemporânea, ainda que pouco explorado. Ainda que os usuários sejam desprovidos de aptidão ou recursos as redes sociais permitem aos usuários organizarem suas ações em prol de socializar e produzir mobilização, pelo próprio desenvolvimento da rede.

A (STATISTICS, 2018a) apresenta um estudo sobre as redes sociais online mais populares em todo o mundo classificadas pelo número de contas ativas até abril de 2018, mostrados na Figura 3.

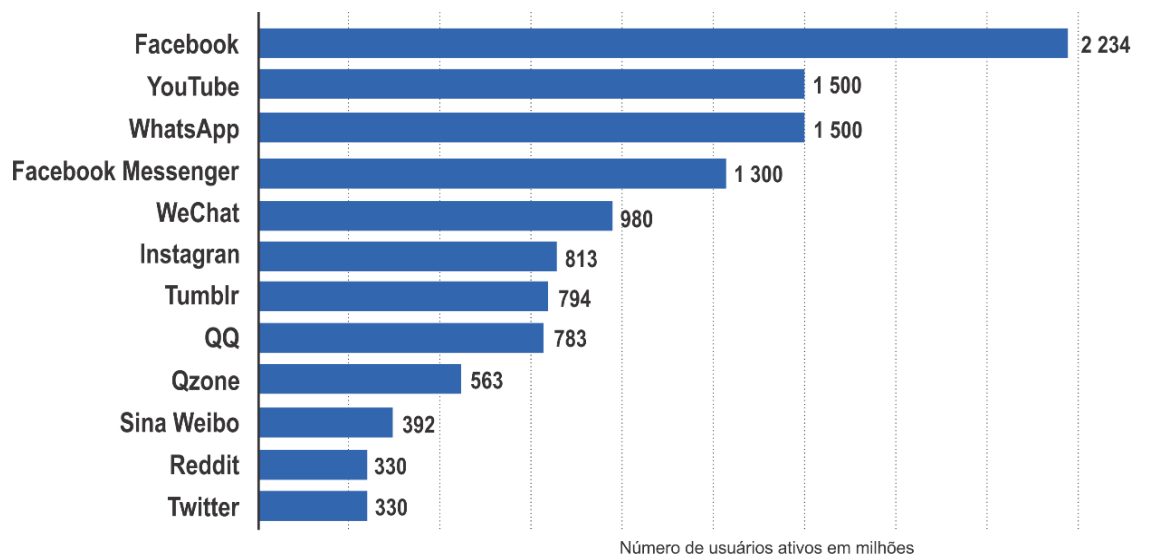


Figura 3. Redes sociais mais populares ao redor do mundo.
Fonte: Adaptado pelo autor (STATISTICS, 2018).

É inegável o impacto que as redes sociais têm sobre a vida de seus usuários, decididamente as transformações ocasionadas na forma com que as pessoas interagem entre si e com o mundo ganharam novos conceitos. Tais aspectos recém surgidos ganham espaço e

abrem discussões sobre os conceitos de identidade digital e interações sociais online, levando a indefinição entre de vida off-line e virtual, criando assim novos paradigmas na sociedade e impactando diretamente nela mesma (STATISTICS, 2018).

2.3 FACEBOOK COMO FERRAMENTA PARA GESTÃO COMPETITIVA

O *Facebook* foi fundado em 2004, pelo então estudante de Harvard Mark Zuckerberg e alguns de seus contemporâneos, inicialmente disponível apenas para estudantes de Harvard, posteriormente expandindo-se para outras universidades regionais, estudantes de ensino médio acima de 13 anos e outros usuários globais. No terceiro trimestre de 2012, o *Facebook* se tornou a primeira rede social a atingir a marca de um bilhão de usuários ativos e atualmente conta com cerca de 2,19 bilhões de usuários ativos mensais. Atualmente o *Facebook* é utilizado por cerca de 80% dos usuários de internet, tornam-se líder no mercado de redes sociais (STATISTICS, 2018a).

Para usar o *Facebook* os usuários precisam de registrar, onde podem criar um perfil pessoal livremente para interagir com outros usuários que adicionem como amigos. O *Facebook* possibilita ainda que os usuários participem de grupos com base no local onde trabalham, estudam ou grupos com os quais possuam interesses sobre determinado assunto, além de poderem criar categorias dentre seus contatos colocando-os em listas. Os usuários podem atualizar seus status através de postagens ou outros conteúdos e enviar mensagens entre si. É possível também interagir com inúmeras aplicações existentes dentro do próprio *Facebook*, como jogos outros serviços e com o Instagram que um aplicativo de compartilhamento de fotos (STATISTICS, 2018a).

O *Facebook* além de permite a criação de perfis pessoais, possibilita às organizações o recurso de *fan pages*. Com elas é possível compartilhar histórias, criar campanhas, divulgar promoções de uma organização. O recurso de *fan pages*, foram criados em 2007, disponibilizando para as organizações recursos adicionais como aplicativos, publicar história e promover eventos através do perfil, além de disponibilizar a funcionalidade de múltiplos administradores. Uma outra funcionalidade das *fan pages* permite que o administrador da página visualize através de painéis de informações a interação das pessoas com a página. (TAVARES; PAES, 2015).

Dados levantados pela (STATISTICS, 2018b), mostram o número de usuários ativos mensais no período de 2008 a 2018 do Facebook apresentados na Figura 4.

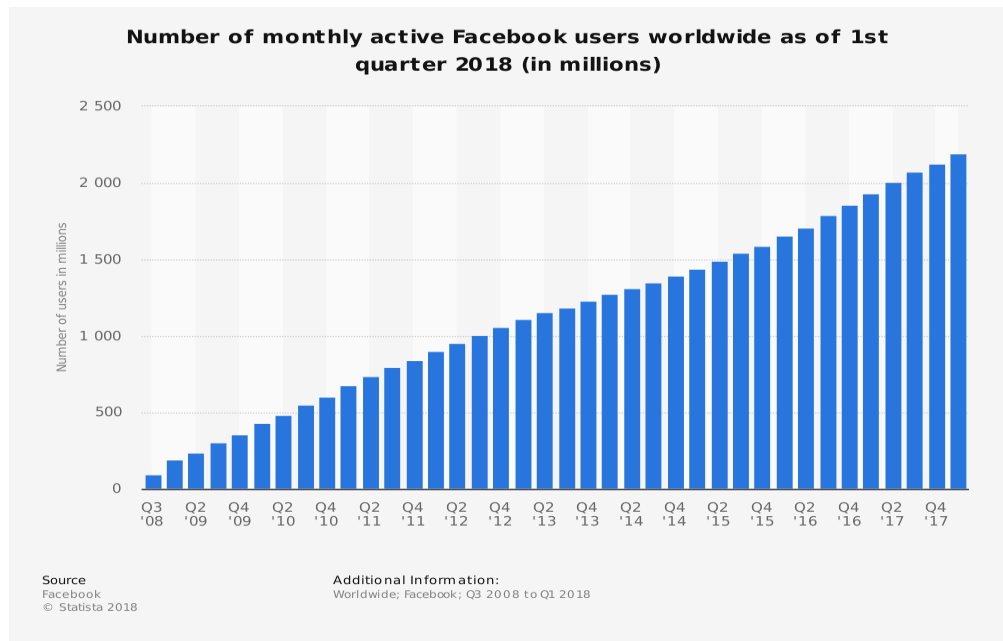


Figura 4. Número de usuários ativos mensais no *Facebook*.
Fonte: (STATISTICS, 2018).

O aumento no número de usuários do Facebook e de outras redes sociais, deve-se muito a popularidade do uso de dispositivos móveis, estima-se que 1,74 bilhões de usuários acessaram suas contas através de plataformas móveis, tornando-a a maior rede social do mundo (STATISTICS, 2018b).

Como o processo de globalização consolidado, o ecossistema empresarial tornou-se extremamente ágil e competitivo. Fato este, obriga as organizações a empregarem cada vez mais tempo e energia na busca por maneiras inovadoras de agregar conhecimento que possibilite a geração de vantagem competitiva pela mesma perante o segmento de mercado no qual está inserida (SILVA; ROBERTO; OLIVEIRA, 2016).

Entretanto torna o processo de gestão mais competitivo, depende de como são gerenciadas as ferramentas para obtenção de conhecimentos acerca do ramo de atuação da organização. A inteligência competitiva é apoiada em dois pilares fundamentais, a informação e a agilidade com que as informações são usadas. Isso porque a inteligência competitiva utiliza-se de tipos e fontes diversas de informação em altas velocidades para se monitorar (BATTAGLIA, 1999).

Todo este dinamismo na geração de inteligência competitiva tornou-se mais agudo através da utilização das tecnologias baseadas na internet como as mídias sociais. Nos últimos anos as mudanças acontecidas na prática organizacional, tanto relacionadas às marcas quanto aos consumidores foram aquecidas pelo uso apropriado da internet, que se converteu em uma

fonte primordial para alcançar produtividade e competitividade independentemente do tipo ou ramo da organização (CASTELLS, 2003).

Dentre as tecnologias de mídias sociais que ganharam cada vez mais notoriedade por propiciar a geração de inteligência de negócio destaca-se as redes sociais online. Devido a interatividade que elas oportunizam, acabaram se transformando em um ambiente que naturalmente é capaz de alargar o relacionamento entre pessoas, bem como a relação clientes e organizações e ainda servir de canal de comunicação capaz de intensificar esses relacionamentos. Este novo paradigma que as redes sociais trouxeram para a vida das empresas abriu uma gama de possibilidades, como fazer novos negócios, interagir com os clientes, comercializar seus produtos e serviços, publicar conteúdo relevante para a marca e realizar campanhas de marketing pela rede. Tais possibilidades mostram o quão expressiva as redes sociais se tornaram na vida de uma organização, apontando uma tendência mundial para gestão organizacional, de cada vez mais empresas buscarem está mais próxima dos consumidor através da internet (BARCELLOS, 2010; SILVA; ROBERTO; OLIVEIRA, 2016).

Uma pesquisa realizada pelas consultorias Centro de Inteligência Padrão (CIP) e *Mind Miners*, que teve parte dos resultados apresentados pela *GloboNews* em uma matéria, mostrou que 56% dos consumidores entrevistados compartilham suas experiências de positivas de compras em suas redes sociais e 54% seguem o perfil das marcas que costumam realizar compras. Além disso, a mesma pesquisa revelou que 21% dos consumidores costumam visitar as páginas da marca em redes sociais como fonte de pesquisa antes de tomarem a decisão final sobre uma compra. Cerca de 80% dos entrevistados afirmam ler os comentários de outros usuários sobre o produto ou serviço oferecido antes de comprar pela primeira vez, destes 71% dizem confiar opinião de outras pessoas que já adquiriram o item oferecido (FRANKENTHAL, 2018).

Ainda segundo Frankenthal (2018), ou mesmo tempo que as redes sociais podem representar grandes oportunidades para as organizações no que tangem a interação, aproximação e possibilitar as marcas conhecerem melhor seu público, bem como suas necessidades, demandas e desejos de consumo, elas representam igualmente um desafio para os gestores tomarem decisões estratégicas para as organizações, uma vez que os consumidores estão mais atentos e informados sobre as organizações. Levando em conta este pensamento as redes sociais acabam se tornando um meio que as organizações têm de dialogar de forma aberta com os consumidores.

2.4 MINERAÇÃO DE DADOS EM REDES SOCIAIS

Para entender o conceito de mineração de dados em redes sociais é preciso primeiramente compreender a mineração de dados em si. Para Camilo e Silva (2009), dada sua aplicabilidade variar de acordo como a área que se quer descobrir novos conhecimentos e as técnicas utilizadas, a mineração de dados pode ser considerada como uma área multidisciplinar, se destacando entre elas dentro da mineração de dados a estatística, aprendizado de máquina e banco de dados. Esta característica multidisciplinar da mineração de dados e o êxito em descobrir novos conhecimentos justifica o crescente número de publicações que a integram juntamente as mais diversas áreas do conhecimento humano.

Analisando a mineração de dados do ponto de vista da estatística, o objetivo de analisar um grande volume de dados é encontrar o conhecimento implícito e torná-lo acessível formatando para que o conhecimento assim gerado seja útil ao possuidor dos dados. Do ponto de vista de banco de dados, a mineração de dados é a junção de técnicas que englobam diferentes contextos de acordo com a necessidade, sendo que as principais técnicas empregadas são aprendizado de máquina, reconhecimento de padrões, estatísticas, banco de dados e visualização, aplicadas em grandes bases de dados para descobrir novos conhecimento. Na ótica de aprendizado de máquina a mineração de dados compõem uma etapa do processo de descoberta de conhecimento que é mostrado na Figura 2, através da aplicação de algoritmos de aprendizagem, classificação ou baseados em estatística a mineração é capaz de extrair ou tornar mais evidentes certos padrões (HAND et al., 2001; CABENA et al., 1998; FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996;).

Embora a mineração de dados apoiasse na utilização de ferramentas bem desenvolvidas que empregam algoritmos altamente eficazes na tarefa de recuperação e análise de conhecimento, a automatização completa do processo de descoberta de conhecimento é algo intangível, uma vez que é necessário a verificação dos resultados, feita por pessoas para avaliar como o processo está sendo executado. Mesmo assim, no processo de KDD a mineração de dados é uma etapa fundamental para obtenção das informações mais relevantes para a tarefa proposta (THOMAZ, 2014).

Na visão de (REZENDE, 2003) a definição mais aceita de mineração de dados foi elaborada por Fayyad, Piatetsky-Shapiro e Smyth, que é a expressão da percepção da área de aprendizagem de máquina, sendo como: “Extração de Conhecimento de Base de Dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados” (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Para ter a compreensão destes conceitos é preciso olhar individualmente para cada item. O Quadro 2 mostra os componentes e conceitos fundamentais para atingir este objetivo listando os componentes e sua definição.

COMPONENTE	DEFINIÇÃO
DADOS	Conjunto de fatos ou casos em um repositório de dados. Por exemplo, os dados correspondem aos valores dos campos de um registro de vendas em uma Base de Dados qualquer;
PADRÕES	Denota alguma abstração de um subconjunto dos dados em alguma linguagem descritiva de conceitos;
PROCESSO	A Extração de Conhecimento de Base de Dados envolve diversas etapas como a preparação dos dados, busca por padrões e avaliação do conhecimento;
VÁLIDOS	Os padrões descobertos devem possuir algum grau de certeza, ou seja, devem satisfazer funções ou limiares que garantam que os exemplos cobertos e os casos relacionados ao padrão encontrado sejam aceitáveis;
NOVOS	Um padrão encontrado deve fornecer novas informações sobre os dados. O grau de novidade serve para determinar quão novo ou inédito é um padrão. Pode ser medido por meio de comparações entre as mudanças ocorridas nos dados ou no conhecimento anterior;
ÚTEIS	Os padrões descobertos devem ser incorporados para serem utilizados;
COMPREENSÍVEIS	Um dos objetivos de realizar mineração de dados é encontrar padrões descritos em alguma linguagem que pode ser compreendida pelos usuários permitindo uma análise mais profunda dos dados;

Quadro 2. Conceitos importantes sobre extração de conhecimento em base de dados.

Fonte: Rezende (2003).

No processo de descoberta de conhecimento em bases de dados é significativo que as informações sejam úteis e entendíveis para os humanos, desta forma o conhecimento adquirido atua como ferramenta de suporte a decisões estratégicas. Além disso é necessário a estruturação dos dados a serem usados uma vez que o processo depende do tipo de dado de entrada para aplicação da técnica de mineração mais adequada (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

A extração de dados em redes sociais exige a aplicação de técnicas diferentes das aplicadas na mineração de dados clássica em base de dados. Dada as características das redes sociais conforme apresentado na Seção 2.2, as pesquisas na área utilizam essencialmente técnicas de mineração para grafos devido a estrutura de entidades e seus relacionamentos assemelhar-se como a estrutura de nodos e arestas presentes nos grafos (CERVI, 2008).

Devido à grande quantidade de informações trafegadas pelas redes sociais, os sistemas de coleta e análise de dados precisam ser potentes o suficiente para suportar a carga de dados que

transitam por elas. Essencialmente a maioria os dados capturados nas redes sociais, são dados não estruturados como fotos, vídeos, compartilhamentos entre outro tipo de postagem. Outra atribuição para esse tipo de aplicação e que além de minerar os dados eles devem ter a capacidade de pesquisar por dados em outros idiomas, tendo em vista que a internet, os comentários sobre negócios, produtos ou marcas podem ocorrer em diferentes partes do mundo (HEKIMA, 2015).

Além do valor que as redes sociais são capazes de agregar no meio corporativo, o aumento da utilização das mídias sociais por indivíduos comuns disponibiliza aos pesquisadores dados sobre o comportamento humano. Quando os usuários interagem com algum tipo de rede social ele deixa um rastro, que pode ser registrado e parametrizado quanto ao que tange a maneira como ele se porta quando está online. Apesar das políticas de privacidade limitarem o acesso a pluralidade dos dados, é possível ter acesso a uma boa parte deles através de interfaces de programação de aplicativo (*Application Programming Interfaces* – APIs) disponíveis em alguns serviços como Google, Twitter, Facebook. Essas interfaces permitem ao pesquisador acessar os dados de publicações e comentários por exemplo do Facebook.

Além das APIs, existem inúmeras outras ferramentas e sites que possibilitam realizar procedimentos de monitoramento e análise em mídias sociais, porém em sua maioria, apresentam de deficiência quando se trata de grandes bases de dados que possuam um cenário personalizado ou que necessite que seja efetuada uma análise mais detalhada sobre os dados. Nesses cenários, o desenvolvimento de ferramentas capazes de realizar a descoberta de conhecimento faz-se necessário ou pelo menos parte do processo para a se alcança a melhor execução da tarefa a partir dos dados estudados (CROOKS et al., 2013).

As próximas seções foram dedicadas a discorrer sobre os modelos de extração e análise de dados encontrados na literatura, tendo em vista que existem outras abordagens para esta tarefa, além de analisar os pontos fortes e fracos de cada metodologia citada, identificando os principais processos que envolvem a construção de uma metodologia eficiente para esta finalidade.

2.4.1 Modelo proposto por Neves (2013)

Para Neves (2013) destaque-se que o uso que as mídias sociais têm em relação a manifestação de opinião pelos consumidores. Isso se deve principalmente pelas características apresentadas pelas mídias sociais que facilitam a comunicação e o diálogo. O Twitter entre

outras vem ganhando cada vez mais notoriedade no contexto corporativo, tornando-se importantes no que tangem a forma de relacionar-se com os consumidores.

A Figura 5 ilustra o processo de opinião do consumidor elaborado pelo autor.

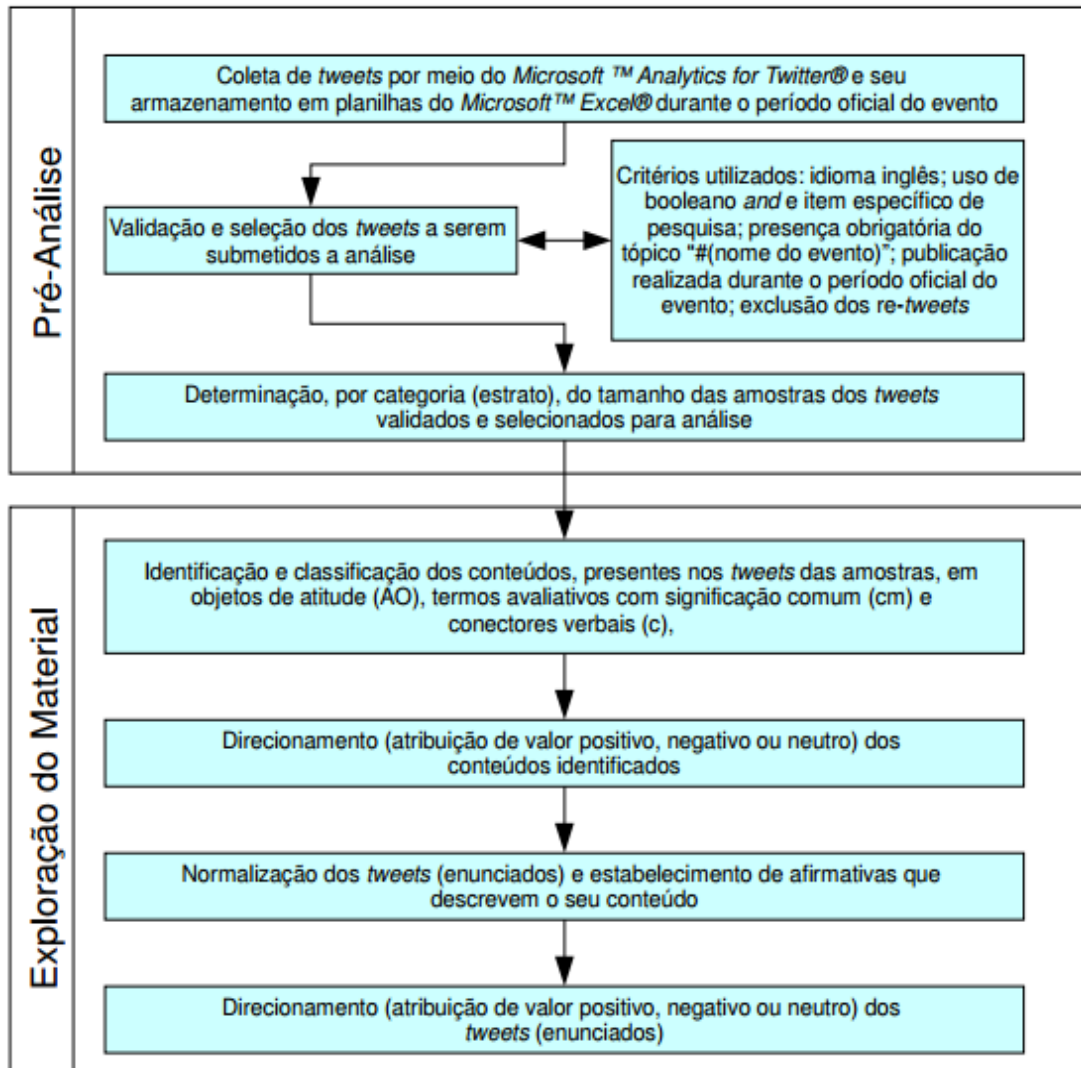


Figura 5. Metodologia de coleta e análise de *tweets* proposta por Neves.

Fonte: Neves (2013, p.70).

A Figura 5 representa as etapas e fases sugeridas pelo autor, os *tweets* recuperados foram armazenados em planilhas do *Microsoft™ Excel®*. O trabalho de Neves (2013) merece destaque pois visou a aplicação prática no setor de turismo levando em consideração a qualidade que os consumidores percebem em relação aos produtos e serviços que foram avaliados em um evento de grande porte, no caso, as Olimpíadas de Londres. Utilizando-se de *hashtags* e busca booleana como estratégia, o autor pode direcionar mensagens e adequar os seus objetivos durante a pesquisa.

O autor discorre em seu trabalho sobre o uso de métodos e ferramentas para analisar as postagens no *Twitter* relacionadas a percepção do consumidor sobre a qualidade de produtos e serviços turísticos oferecidos durante os jogos Olímpicos de Londres 2012. Neves utilizou-se do aplicativo da *Microsoft Analytics for Twitter* para armazenar os tweets que foram analisados de acordo com as categorias identificadas pelo autor para indicar a opinião do consumidor sobre os serviços turísticos relacionados ao evento, dentre elas: alimentação, hospedagem, transporte e segurança (THOMAZ, 2014).

2.4.2 Modelo proposto por Hea, Zha e Li (2013)

A necessidade das empresas de avaliar o ambiente competitivo na qual estão inseridas é fundamental para o sucesso de uma organização. É percebendo essa necessidade que muitas organizações estão adotando as redes sociais como ferramenta para se comunicar com os consumidores. O resultado disto é uma grande quantidade de dados gerados através desta interação nas redes sociais com *Facebook* e *Twitter*, mas para transformar estes dados em inteligência competitiva e preciso monitorar e analisar não somente os dados que os clientes da organização gera, mais os dados gerados nas redes sociais dos concorrentes (HEA; ZHA; LI, 2013).

A Figura 6 mostra as etapas metodológicas adotadas por Hea, Zha e Li (2013).

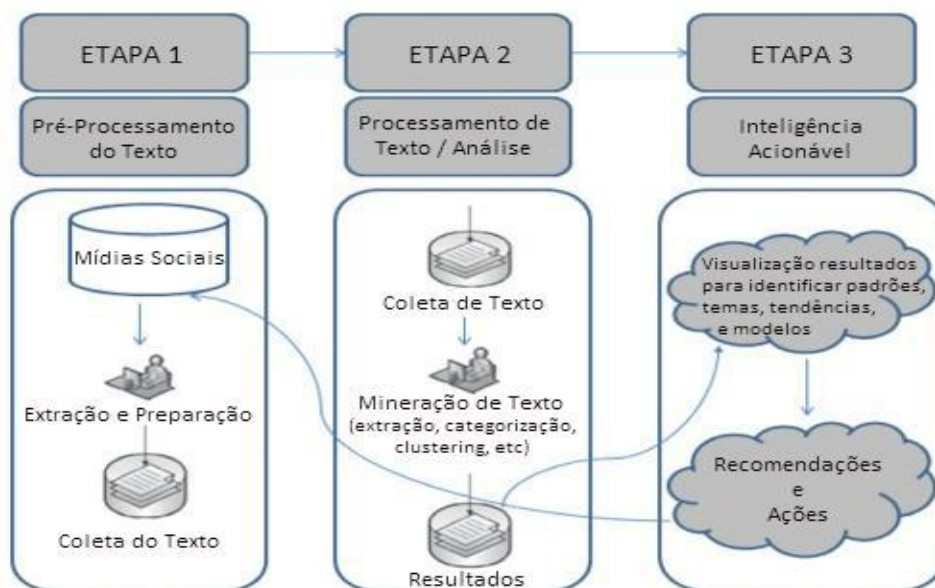


Figura 6. Metodologia de mineração conteúdo texto em mídias sociais.

Fonte: Hea, Zha e Li (2013, p.3).

Conforme mostra a Figura 6, a metodologia foi composta por 3 etapas: a primeira consistiu na coleta dos dados quantitativos individualmente nas páginas no *Facebook* e nos perfis no *Twitter* das organizações como: número de fãs, seguidores, número de postagens, comentários, ações, gostos e frequência das publicações. Em seguida, foram coletados textos e aplicou-se técnicas de mineração de textos visando encontrar novos conhecimentos e padrões analisando as mensagens e conteúdos publicados pelas pizzarias. Por fim, os resultados obtidos foram analisados e apresentados os padrões descobertos bem com a recomendação de ações as organizações.

O modelo foi pensando em como auxiliar as organizações a entender como realizar uma análise em redes sociais para gerar inteligência competitiva para o negócio, transformando os dados em informações que Hea, Zha e Li (2013), elaboraram seu estudo para auxiliar os gestores no processo de tomada de decisão. Os autores propuseram seu modelo e aplicaram técnicas de mineração de texto para analisar textos não estruturados no *Facebook* e no *Twitter* nas páginas das três maiores redes de pizzarias dos Estados Unidos, sendo estas: Pizza Hut, Domino's Pizza e Papa John's Pizza.

2.4.3 Modelo proposto por Abrahams et al., (2013)

Usando o setor automotivo como amostra, Abrahams et al., (2013) implementaram e ajustaram os parâmetros de uma modelo de mineração de texto em grandes volumes de dados em mídias sociais. O modelo proposto segundo os autores seria capaz de isolar automaticamente os componentes e ou equipamentos em fóruns de discussão de usuários com precisão. O processo também reconhece rapidamente diferentes termos para cada categoria de componente, gerando inteligência competitiva e de marketing, nos diferentes níveis da cadeia produtiva do setor.

O modelo apresenta oito fases de implementação tem início a partir do momento que o usuário acessa e publica em um fórum de discussão sobre automóveis. A aquisição dos termos e feita por uma *web crawler* (rastreador web), realiza a busca nos bancos de dados dos fóruns, para a realização da mineração de texto. Fazendo uma analogia a publicações em redes sociais, a primeira fazer seria o equivalente a quando um usuário do Facebook ou Twitter publica algum conteúdo em uma página, entre outro.

A Figura 7 apresenta o modelo proposto por Abrahams et al. (2013).

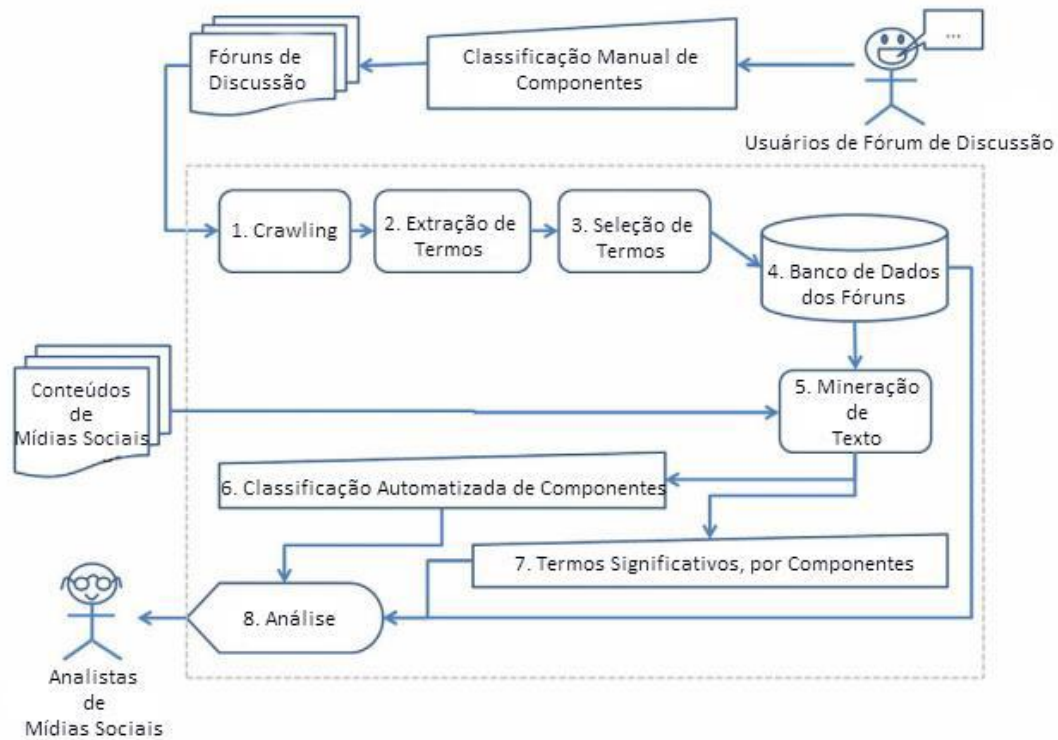


Figura 7. Processo de isolamento de componentes de veículos em mídias sociais.

Fonte: Abrahams et al. (2013).

A metodologia desenvolvida por Abrahams et al. (2013), a primeira fase corresponde à etapa de coleta dos dados nos fóruns de discussão. A segunda consiste na extração dos termos decorrentes dos textos. A terceira fase consiste na seleção dos termos. A quarta fase consiste no processo de armazenamento dos termos selecionados na terceira fase em uma base de dados consistente. Na quinta são realizados os processamentos em cima dos termos a fim de indicar a categorização dos termos. A Sexta fase consiste na classificação automática dos termos. A sétima fase e a geração da lista de termos mais expressivos por categoria. E a oitava fase consiste em descrever analiticamente dos o processo (THOMAZ, 2014).

A metodologia desenvolvida por (ABRAHAMS et al., 2013) destaca-se pela versatilidade permitindo a replicação em outras mídias sociais, e por ser fundamentada em estudos realizados anteriormente e pesquisas consistentes.

2.4.4 Modelo proposto por Thomaz (2014)

Com o objetivo de auxiliar as organizações governamentais, administrativas e organizações privadas do setor de turismo, o autor elaborou seu estudo é embasado na

metodologia desenvolvida por outros autores, propôs um processo de mineração de conteúdo em mídias sociais voltada para auxiliar na gestão de destinos turísticos compostos por sete fases. O processo desenvolvido pelo autor apresentou-se eficiente no aspecto da coleta de conteúdo e satisfatório ao identificar assuntos populares nas mídias sociais realizando análises quantitativas e qualitativas.

Através de seu estudo que foi aplicado na cidade de Curitiba, verificou que nenhuma organização responsável pelo gerenciamento de destinos turísticos utilizava ferramentas para monitorar as mídias sociais ou quaisquer técnicas de mineração de conteúdo. Embora que superficialmente, as organizações de gerenciamento de destinos turísticos faziam o uso da ferramenta integrada analítica do facebook para verificar o nível de desempenhos atingidos pelas ações publicadas nas mídias sociais. O autor completa que, a pouco ou nenhuma utilização de tais ferramentas é atribuída a inexperiência, falta de conhecimento técnico e falta de recursos humanos e/ou financeiros para justificar a não utilização do monitoramento de mídias sociais.

A Figura 8 ilustra os métodos propostos pelo autor para a realização da descoberta de conhecimento em mídias sociais voltadas ao turismo.

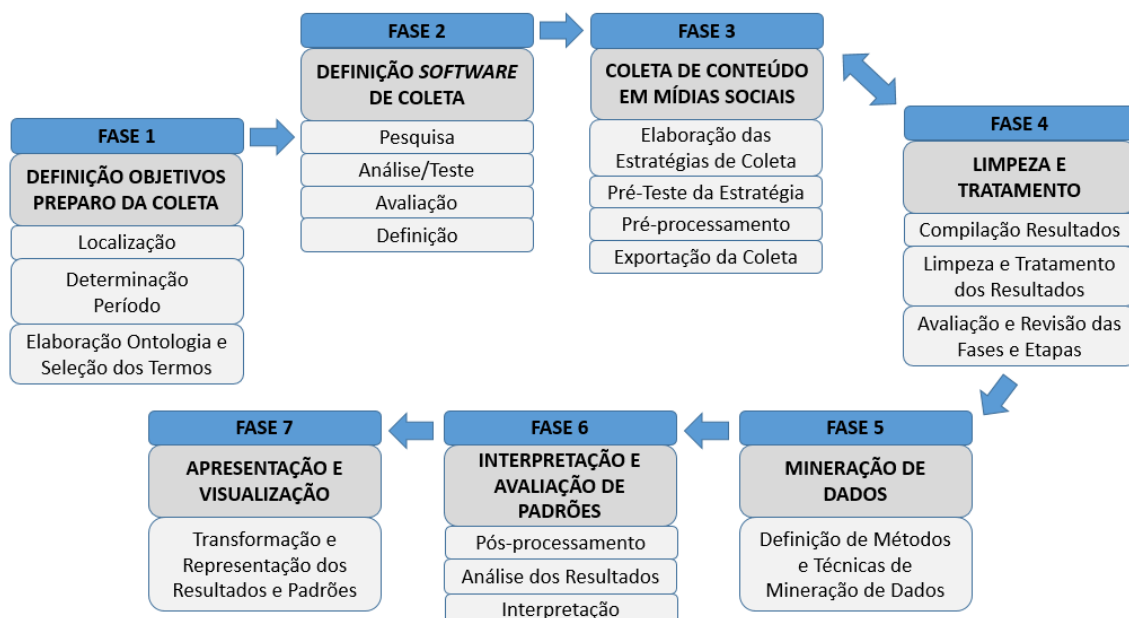


Figura 8. Processo de mineração de conteúdo em mídias sociais para auxiliar na gestão de destinos turísticos.

Fonte: (THOMAZ, 2014).

Em seu estudo Thomaz (2014) agregou as estruturas propostas por (NEVES, 2013), (HEA; ZHA; LI, 2013), (KALAMPOKIS; TAMBOURIS; TARABANIS, 2013), (ABRAHAMS et al., 2013) e incluiu modelos de descoberta de conhecimento apresentados por (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), (CHAPMAN et al., 2000), (HAN;

KAMBER; PEI, 2012). Além disto, Thomaz (2014) examina o monitoramento em mídias sociais através das redes sociais *Facebook*, *Twitter* e *YouTuber* verificando os procedimentos metodológicos adotados no *Twitter* através de mineração de conteúdos ligados a ontologia de aplicação de atrativos e serviços turísticos.

2.4.5 Modelo proposto por Santos (2014)

O trabalho de Santos (2014) buscou identificar meios para coletar e analisar dados nas *fan pages* da maior rede social do mundo na atualidade, a autora através de revisão da literatura investigou os mecanismos em formato de *scripts* que realizassem tal tarefa de forma a satisfazer a etapa de extração de dados proposta no estudo. Utilizando os instrumentos para a coleta a autora explora os aspectos da rede social *Facebook* contextualizando o uso de tais ferramentas sobre a perspectiva de marketing de relacionamento em algumas das maiores marcas em número de fãs e seguidores.

O processo composto pela autora definiu-se em buscar na literatura conceitos para esturrar e dá suporte durante o desenvolvimento do estudo onde buscou-se conhecer o ambiente de explorado pela pesquisa, no caso, as *fan pages* da rede social *Facebook*. A etapa seguinte buscou identificar os artefatos tecnológicos a serem utilizados durante a etapa de coleta de dados, onde definiu-se os *scripts* bases para a coleta. Posteriormente a autora modificou os *scripts* a ser usado para coleta de dados a fim de adequar melhor a seus objetivos. Seguindo a metodologia estabelecida a autora realizou um estudo exploratório onde foram coletados dados da *fan pages* da marca “O BOTICÁRIO”.

Após a coleta exploratória a autora realizar a coleta propriamente dita, utilizando-se de dois instrumentos: o primeiro corresponde a observação sistemática da *fan pages* da marca; o segundo, uma entrevista com os gestores de marketing da marca. Nessa etapa foram estudados a marca “O BOTICÁRIO” e “NATURA”, a escolha baseou-se por se tratarem de duas marcas entre a 100 maiores no número de curtidas e por atuarem no mesmo seguimento. A análise dos dados foi realizada usando duas ferramentas: *Microsoft NodeXL*, para a visualização de gráficos de rede e obtenção das métricas; *Wordle*, aplicativo *online* para a geração de nuvem de palavra.

3 MATERIAIS E MÉTODOS DA PESQUISA

3.1 CARACTERIZAÇÃO DA PESQUISA

Para Vergara (1997), uma pesquisa pode ser rotulada seguindo cinco preceitos: no que concerne ao propósito, à natureza dos dados, ao esboço, ao método de aquisição dos dados e quanto a essência do estudo. O Quadro 3 apresenta as características da presente pesquisa alinhados com os critérios definidos pela autora.

CRITÉRIO	NATUREZA	DEFINIÇÃO
Propósito	Explicativa	Analisar as diferentes técnicas de mineração de dados em mídias sociais, e estabelece uma metodologia para extração e análise de dados em <i>fan pages</i> do <i>Facebook</i> .
Natureza dos dados	Mista	Utiliza de dados qualitativos (textos postados por usuários) e dados quantitativos (estatísticas das páginas, número de seguidores, número de curtidas, reações dos usuários número de comentários por postagem).
Delineamento	Ex-post-facto	Investigação metódica e empírica na qual o pesquisador não possui o controle sobre as variáveis, pois já ocorreram as variações nas variáveis.
Método de aquisição dos dados	Redes sociais	Os dados utilizados na pesquisa foram coletados através de experimentos redes sociais, mais especificamente em <i>fan pages</i> do <i>Facebook</i>
Essência do estudo	Aplicada	Possibilita a geração de conhecimento visando aplicação prática, disponibilizando aos interessados uma metodologia para analisar o efeito das postagens nas <i>fan pages</i> , aceitação de produtos e percepção do consumidor com relação aos concorrentes.

Quadro 3. Características da pesquisa quanto ao propósito natureza dos dados delineamento e natureza da pesquisa.

Fonte: Próprio autor (2018).

3.2 AMBIENTE DE PESQUISA

O desenvolvimento desta pesquisa se deu no contexto das redes sociais, mais especificamente nas *fan pages* da maior rede social do mundo na atualidade, o *Facebook*. Esta mídia permite aos seus usuários criarem perfis pessoais ou no caso de uma marca, existe a possibilidade de criar páginas comerciais. Este recurso de páginas comerciais ou *fan pages* torna-se interessante para as organizações uma vez que permite que organizações tenham um canal direto de comunicação com os fãs da marca, além de possibilitar a interação com

potenciais clientes ou até mesmo realizem negociações, promoção de produtos ou eventos além campanhas de *Marketing*.

Outros recursos interessantes para as organizações, é a possibilidade de criar abas personalizadas com conteúdo próprio da organização como fotos, vídeos, fóruns de discussão, informações sobre produtos específicos, eventos e avaliações de usuários. É possível criar o aplicativo próprio da empresa e adicionar outras funcionalidades ou recursos que não são disponibilizados. É possível para os fãs e além de outras páginas, contribuírem para que a uma página de uma marca se torne mais conhecida ganhando mais seguidores através de marcações que podem ser feitas em praticamente todos os tipos de atualizações pelo *Facebook* (SIQUEIRA, 2017).

Os dados foram coletados através do aplicativo NETVIZZ disponível no próprio Facebook, que executa consultas sobre a *Graph API* (Interface de Programação de Aplicativos – API do Facebook). A *Graph API* atualmente encontra-se na versão 3.0, ela é a principal maneira dos aplicativos lerem e gravarem no gráfico social do Facebook. De alguma maneira todos os SDKs (*Software Development Kit*, no português: Kit de desenvolvimento de *software*) e produtos presentes no *Facebook* inter-relacionam-se com a *Graph API*, além do que as outras APIs que existem no *Facebook* são extensões dela (FACEBOOK, 2018).

3.3 PROCEDIMENTOS METODOLÓGICOS

A Figura 9, ilustra as etapas do método utilizados pelo autor durante a pesquisa para atingir os objetivos.

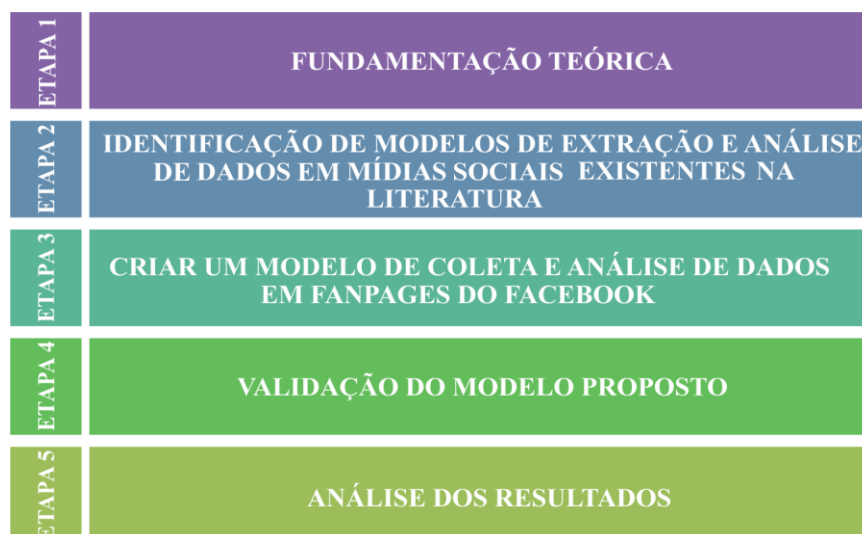


Figura 9. Metodologia da pesquisa.
Fonte: Próprio autor (2018).

Os procedimentos adotados neste estudo, visam atingir os objetivos específicos que serão apresentados no decorrer deste estudo, foi composto por cinco etapas onde buscou-se dividir as tarefas entre pesquisa e desenvolvimento.

Adotou-se a técnica de pesquisa bibliográfica durante a primeira etapa do trabalho, onde buscou-se levantar informações acerca da temática abordada por esta pesquisa. Segundo (DENCKER, 2001), a etapa de levantamento e revisão da literatura é fundamental em toda e qualquer pesquisa, pois trata-se de uma fase crucial para concepção de conceitos e construção do embasamento teórico, devendo assim fazer parte da fase preliminar da pesquisa.

A segunda etapa consistiu na identificação de modelos de coleta e análise de dados em mídias sociais verificando os modelos encontrados pela etapa de fundamentação teórica em artigos, monografias, dissertações e teses. Nesta etapa buscou-se analisar os processos ou parte de processos, observando suas características, procedimentos e metodologias desenvolvidas nos trabalhos encontrados e apresentados nas seções 2.4.1, 2.4.2, 2.4.3, 2.4.4 e 2.4.5. Os modelos citados abordam os processos envolvidos na descoberta de conhecimento a partir das mídias sociais, além de técnicas de extração e análise de dados e aspectos de tangem as ferramentas utilizadas durante o processo, sendo este o ponto de partida para elaboração do modelo proposto.

A terceira etapa, definiu-se a proposta de um modelo prático para coleta e análise de dados em *fan pages* do *Facebook*, com base nos processos identificados pela etapa anterior, buscou-se definir técnicas, métodos e ferramentas preferencialmente as de licença gratuita para executar a metodologia proposto por esta pesquisa. Na quarta etapa deste estudo buscou-se verificar a sistemática apresentada por meio da aplicação na prática dos processos. Essa etapa também permitiu avaliar cada fase da metodologia proposta, é aperfeiçoar os pontos onde era passível de melhorias.

Na quinta e última etapa analisou-se os resultados obtidos, bem como se os objetivos específicos foram atingidos. Nesta etapa se investigou a contribuição que a metodologia proposta deixou para a área na qual foi feita a validação explanando os conhecimentos descobertos nas *fan pages* escolhidas.

3.4 CONSTRUÇÃO DO MODELO DE EXTRAÇÃO E ANÁLISE DE DADOS EM *FAN PAGES*

Para o desenvolvimento da metodologia de extração e análise de dados em *fan pages* do *Facebook*, buscou-se na literatura trabalhos que pudessem auxiliar na construção de um

modelo próprio. O Quadro 4 mostra as principais contribuições dos modelos de extração e análise de dados observados na literatura deram para elaboração da presente pesquisa.

AUTORES	PRINCIPAIS CONTRIBUIÇÃO
Neves (2013)	Diagrama de processos; Fases de coleta, validação e seleção de conteúdo; Análise de opinião do consumidor;
HEA, ZHA E LI (2013)	Pré-Processamento de Texto (Extração, Preparação e Coleta de Conteúdos); Visualização dos resultados para identificar padrões.
ABRAHAMS, JIAO, FAN, WANG E ZHANG (2013)	Filtragem de Dados de Mídias Sociais; Análise de textos.
THOMAZ (2014)	Monitoramento em mídias sociais através das redes sociais <i>Facebook</i> , <i>Twitter</i> e <i>YouTube</i> ; Aplicação no setor de serviços ligados ao turísticos.
SANTOS (2014)	Coletar e analisar dados nas <i>fan pages</i> ; <i>Facebook</i> e o marketing de relacionamento; Ferramentas de coleta de dados <i>fan pages</i> .

Quadro 4. Contribuição dos autores para elaboração do método de extração e análise em *fan pages* do *Facebook*.

Fonte: Próprio autor (2018).

Por meio da verificação dos principais processos envolvidos na descoberta de conhecimento em mídias sociais, foi possível elaborar um método que englobasse as principais técnicas envolvidas na extração e análise dos dados na rede social *Facebook*. A construção do processo deu-se de tal forma a envolver processos simples, mas eficazes, utilizando ferramentas gratuitas e *open source* (código aberto), minimizando os custos de implementação. Todo o processo dividiu-se em quatro etapas, sendo estas: etapa de coleta, etapa de pré-processamento, etapa de processamento e etapa de análise e visualização. Cada etapa apresenta em seu escopo processos complementares.

A primeira etapa de coleta diz respeito aquisição dos dados, utilizando técnicas específicas para o tipo de dado ou ambiente no qual se deseja fazer a aquisição dos dados. Na etapa de pré-processamento e feita a verificação dos dados coletados, bem como a avaliação da coleta, onde os dados inconsistentes ou inválidos serão descartados para poder construir uma base sólida. O processamento dos dados é a terceira etapa da metodologia, e foi dividido em três subprocessos: mineração de texto; mineração de dados e Tratamento da rede de relacionamento da página. Esta etapa foi pensada para o melhor aproveitamento das informações recuperadas. A quarta e última etapa de análise e visualização visou verificar os resultados alcançados

identificando se os objetivos de descoberta de conhecimento foram efetivados e apresentando os resultados através de representações gráficas para o melhor entendimento dos mesmos.

O Quadro 5 apresenta as etapas citadas e os processos de cada etapa, bem como sua descrição.

ETAPAS	PROCESSO	DESCRIÇÃO
Coleta dos Dados	Definição do Ambiente de Coleta	Verificar e definir quais <i>fan pages</i> deveram passar pelo processo de descoberta de conhecimento.
	Definição das Estratégias de Coleta	Definição dos procedimentos para a coleta de dados e avaliação.
	Coleta de Dados	Extração de dados utilizando a aplicação Netvizz.
Pré-Processamento	Verificação, Seleção e Validação dos Dados Coletados	Verificação das consistências dos dados, limpeza e remoção de valores inválidos e aprovação dos dados.
Processamento	Mineração de texto	Classificação automatizada de postagens por meio da análise de sentimentos.
	Mineração de dado	Mineração de regras de associação para identificar padrões nos dados estáticos das postagens utilizando a ferramenta <i>Weka</i> e o algoritmo <i>apriori</i> .
	Tratamento Da rede de relacionamento da página	Formatação dos dados contidos nas redes para gerar resultados legíveis para análise visual.
Análise e Visualização dos Resultados	Análise dos resultados	Verificação dos resultados alcançado na etapa de processamento.
	Visualização	Mostrar os resultados alcançados, conclusões, nuvem de palavras e grafo de relacionamento.

Quadro 5. Etapas e processos da metodologia de extração e análise de *fan pages*.

Fonte: Próprio autor (2018).

A análise das metodologias foi essencial para a obtenção da compreensão de cada etapa, fase e processos observados. Foram combinadas diferentes técnicas de descoberta de conhecimento que resultaram em uma metodologia de extração e análise de dados no *Facebook* com foco nas *fan pages*.

A aplicação do modelo proposto começa na etapa de coleta, onde pode-se iniciar pelo processo de definição de estratégias de coleta ou definição do ambiente de coleta. Quando o ambiente foi previamente definido, então o início é a definição da estratégia de coleta, caso contrário, define-se primeiro ambiente de coleta e depois a estratégia, e pôr fim a coleta propriamente dita.

Figura 10 apresenta as etapas do modelo elaborado por este estudo, bem como o seu fluxo de execução.

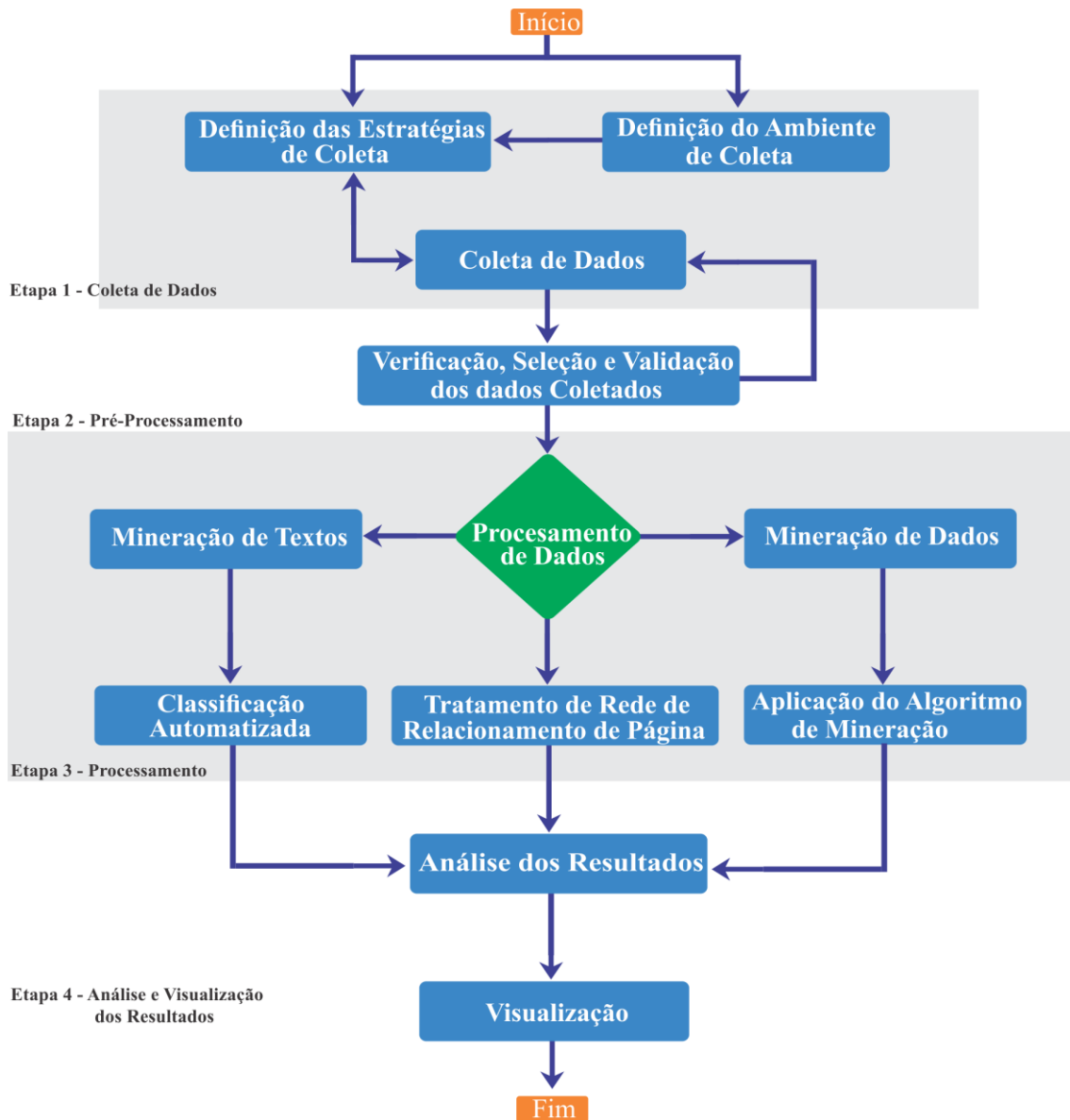


Figura 10. Metodologia proposta para extração e análise de dados em *fan pages* do *Facebook*.

Fonte: Próprio autor (2018).

Após a aquisição dos dados, começa o processo de pré-processamento onde deve-se eliminar todos os dados que apresentem inconsistências, instâncias vazias ou para a mineração de texto todos os que apresentem caracteres inválidos, pois o script primeiro verifica as palavras em português antes de passá-las para a língua inglesa, e caso exista alguma inconsistência pode-se acabar gerando a falha na leitura pelo script.

3.5 FERRAMENTAS

3.5.1 Python E Textblod

O Python é uma linguagem de programação desenvolvida em 1990 no Instituto Nacional de Pesquisa para Matemática e Ciência da Computação da Holanda (CWI), criada por Guido Van Rossum, inicialmente focado em profissionais das áreas de física e engenharia. Seu desenvolvimento se deu a partir de uma outra linguagem de programação chamada ABC, e atualmente se tornou uma linguagem bem aceita no tanto no meio acadêmico quanto na indústria, sendo utilizada por grandes empresas que trabalham com alta tecnologia como: Google, Yahoo, Microsoft entre outras (BORGES, 2010).

O Python atualmente se encontra na versão 3.7, a codificação oficial é mantida pela PSF (*Python Software Foundation*). Além disso a linguagem é escrita em C, é por esta razão também é chamada de CPython. Inclui inúmeras estruturas de alto nível, possuindo uma gama de módulos prontos e diversos *frameworks* produzidos por terceiros que podem ser adicionados (PYTHON, 2018).

A *TextBlod* é uma biblioteca para o Python (2 e 3), focada na manipulação de dados não estruturados fornecendo de maneira simples uma poderosa API para execução de atividades relacionadas ao processamento de linguagem natural (NLP), tais como delimitação de fala, classificação, análise de sentimento, tradução entre outros recursos. Seus recursos são baseados em outras reconhecidas bibliotecas que trabalham com NLP, como as bibliotecas *NLTK* e *Pattern* e também trabalha bem com ambas (TEXTBLOD, 2018).

Existe uma propriedade dentro da análise de sentimento no *TextBlod* chamada *sentiment*, ela será explorada por este trabalho, pois com ela é possível retornar uma pontuação de Polaridade que varia de no intervalo de [-1.0 a 1.0] onde os valores abaixo de 0.0 indicam que o sentimento é negativo, valores acima de 0.0 indicam sentimento positivo e o valor 0.0 significa dizer que é neutro. Além da Polaridade, também é possível medir a Subjetividade de uma palavra ou texto, onde os valores retornados variam de [0.0 a 1.0], indicando assim a porcentagem de subjetividade apresentada no texto onde 1.0 significa 100% de subjetividade (TEXTBLOD, 2018).

3.5.2 Netvizz

O Netvizz é um aplicativo desenvolvido para atuar na plataforma do *Facebook*, com ele é possível extrair dados públicos em perfis de usuários, páginas e grupos específicos para serem utilizados em pesquisas. Ele conta com recursos de recuperação de informações de

grupos, páginas e quantidades de curtidas, além de postagens feitas na linha do tempo, rede de página e estatística de link bastando apenas ao usuário escolher o serviço e fornecer o ID do *Facebook* para iniciar a pesquisa. O aplicativo retorna os resultados público em formato *tab*, uma vez que os dados marcados como privado são protegidos pelas políticas de privacidade do *Facebook* (RIBEIRO, 2018).

O acesso aos dados é feito através da API do *Facebook*, por meio de requisições feitas diretamente a ela. O aplicativo foi escrito na linguagem de programação PHP sendo executado em um servidor que foi fornecido pela *Digital Methods Initiative* (grupo de pesquisa europeu de estudos da Internet). Para poder utilizar o aplicativo basta apenas ou usuário pesquisar na barra de pesquisa do pelo aplicativo e possuir uma conta como usuário do *Facebook* para poder acessar os dados (RIEDER, 2013).

3.5.3 Tagul

A ferramenta Tagul é uma plataforma online voltada para análise e visualização *tag cloud*, ou seja, dados textuais em forma de nuvem de palavras. A apresentação do e feita através de dados estatísticos que se relacionam de acordo com a frequência em que as palavras ou frases mais aparecem em um texto. A ferramenta é disponibilizada de forma gratuita e as nuvens geradas são personalizadas de acordo com as especificações que o usuário definir, oferecendo meios para edição de texto, fontes, cores, *layout*, além de contar com várias formas pré-definidas de nuvem. Para utilizar a ferramenta basta fazer um simples cadastro rápido que pode ser feito usando as redes sociais, como *Facebook* e *Twitter* ou por meio de uma conta no Google, ao final é possível fazer o download da nuvem resultante diretamente no computador (MANNARA, 2016; SANTOS; PRÓSPERO; GIANORDOLI, 2014).

3.5.4 Weka

Waikato Environment for Knowledge Analysis – WEKA é uma aplicação que reuni uma gama de algoritmos de aprendizado de máquina voltados para a descoberta de conhecimento por meio da mineração de dados. As técnicas presentes no *WEKA* podem ser aplicadas diretamente na interface próprio da ferramenta ou podem ser chamadas em outra aplicação e aplicadas diretamente em um conjunto de dados bastando apenas adicionar o *WEKA* nas aplicações de código fonte Java. As ferramentas que o *WEKA* contempla incluem pré-

processamento de dados, classificação, regressão, agrupamento, regras de associação e visualização (*WEKA*, 2018).

O *WEKA* foi desenvolvido na Universidade de Waikato, na Nova Zelândia, no ano de 1993, e seu nome deriva de uma ave que abita a região. O software apresenta uma interface intuitiva que permite interação com dados produzindo, além resultados visuais em formato de tabelas e curvas, de forma que a tela inicial permite a seleção de quatro modos diferentes de utilização do *WEKA* (HALL et al., 2009).

3.5.5 Gephi

O pacote de análise e visualização de redes *Gephi*, foi escrito utilizando a plataforma NetBeans. Trata-se de uma aplicação de código aberto que permite a importa, visualizar, filtrar, navegar e agrupar dados (*clustering*). Sua utilização possibilita descoberta de padrões intuitivamente através de análise visual, além de disponibilizar diversos algoritmos que produzem *layouts* valiosos para visualização do conhecimento. Tem sido usado em vários projetos de pesquisa na academia, jornalismo e em outros lugares, por exemplo, visualizando a conectividade global do conteúdo do *New York Times* e examinando o tráfego da rede *Twitter* (BASTIAN; HEYMANN, 2009).

4 PROPOSTA DE APLICAÇÃO: FACEBOOK NA GESTÃO DO TURISMO NA REGIÃO DO LITORAL PIAUIENSE

E inegável o papel exercido pela internet nos tempos atuais, fazendo com que ela atue como uma importante ferramenta de comunicação e informação para os mais diversos setores da economia. No setor de turismo não é diferente, a interação com a internet ocasiona uma mudança no comportamento do turista, atuando como aporte na hora de escolher um roteiro turístico. Neste aspecto o uso da internet como ferramenta de divulgação do turismo no estado do Piauí agrega grande valor ao setor em todo o estado que possui diversas fontes de recursos naturais (CRUZ et al., 2012).

Entre as diversas riquezas naturais que o estado tem a oferecer a quem o visita, se destaca a região norte com seu litoral. Embora o estado possua a menor costa dentre os outros e está localizado entre os estados do Ceará e Maranhão que vocacionalmente se destacam pelo turismo de sol e praia. O norte piauiense não decepciona ao turista que escolhe os seus 66 quilômetros de faixa costeira como destino turístico, possuindo praias relativamente calmas e belas, além do seu principal atrativo turístico, o delta do Parnaíba (CRUZ et al., 2012; PUREVIAGEM, 2016).

Mesmo com inúmeras riquezas naturais, o turismo no litoral ainda é pouco explorado em comparação por exemplo, com os estados vizinhos. Além da necessidade de investimentos e incentivos advindos do setor público, é latente a insuficiência em termos de divulgação de tais atrativos do estado. Neste sentido, as redes sociais vêm avançando cada vez mais como estratégia de divulgação e disseminação de informações acerca do turismo, principalmente no setor privado. Dentre as redes sociais que merecem destaque o *Facebook*, que pode atuar como ferramenta para implantação de estratégias de divulgação.

Alves (2011) expressa que o *Facebook*, possibilita ao consumidor um não somente uma maneira de realizar consultas ou reclamações sobre uma marca, mais que também pode ser utilizado por elas como ferramenta para realizar promoção eficiente, interação com o público e divulgação de conteúdo relacionados ao turismo. Essa característica demonstra como uma importante possibilidade de estratégia para o setor no estado do Piauí em especial para no seu litoral.

Para validar a metodologia apresentada na seção 3.4 por esta pesquisa, esta seção buscou demonstrar os passos percorridos em cada etapa do modelo elaborado, demonstrando a execução de cada aplicação que foi utilizada para a descoberta de conhecimento em *fan pages* do *Facebook*.

Em sintonia com o modelo proposto por Thomaz (2014), que realizou seu estudo voltado para auxiliar as organizações envolvidas com o setor turístico na cidade de Curitiba, e levando em consideração que, o litoral piauiense riquíssimo em belezas naturais sendo a porta de entrada para o delta do Parnaíba, a realização da aquisição dos dados foi feita com o foco no setor de turismo no litoral piauiense.

A Quadro 6 mostra a relação das agências de turismo presente na região norte do estado, onde primeiro buscou-se fazer um levantamento das organizações presentes na região por meio de pesquisas na internet, e posteriormente com o nome das empresas que atuam na região, procurou-se pelas respectivas páginas no *Facebook*, para levantar os dados referentes às *fan pages* como: número de seguidores curtidas, avaliação, ID do *Facebook* além do site oficial quando esta possuir.

ORGANIZAÇÃO	LIKES	SEGUIDORES	AVALIAÇÃO	ID DO FACEBOOK
Delta Rio Parnaíba Turismo	9.806	9.788	4,7	247237165344508
Natur Turismo	5.635	5.627	Sem avaliações	155109624516180
Rastro Nordestino	7.763	7.739	Sem avaliações	198648356935742
Eco Adventure	160.118	160.120	4,8	147651511929605
Clip Ecoturismo e Aventure	14.233	14.230	Sem avaliações	297111820335738
Atalaia Turismo Ltda.	848	831	Sem avaliações	159904214176891
Sun Road Turismo	3.572	3.565	Sem avaliações	700705556718384
Acqua Turismo	5.450	5.492	4,8	752424301516665
CVC Parnaíba	10.339	10.412	4,9	919279878193722

Quadro 6. Maiores agências de turismo em número de seguidores no litoral piauiense.

Fonte: Próprio autor (2018).

Para etapa de coleta foram selecionadas 3 agências, sendo elas: Eco Adventure, Clip ecoturismo e aventure e CVC Parnaíba. A escolha se baseou no número de *likes* e no número de seguidores.

4.1 ETAPA 1 – COLETA DE DADOS

Na etapa um, após definir as *fan pages* a serem analisadas foi realizado a coleta de dados no aplicativo Netvizz. Foram explorados dois módulos diferentes nesta etapa da pesquisa, sendo o primeiro relativo ao *Page data* (Dados de página), que criam arquivos de tabela para as atividades dos usuários em relação às postagens feitas nas páginas. O segundo módulo

explorado, diz respeito ao *Page like network* (Página como rede), cria uma rede de relacionamento de páginas conectadas através *likes* entre elas (RIEDER, 2013).

A ferramenta apresenta uma interface simples e intuitiva de fácil interação, a figura 11, mostra a interface do aplicativo que logo na tela inicial apresenta todos os módulos que a aplicação possui.

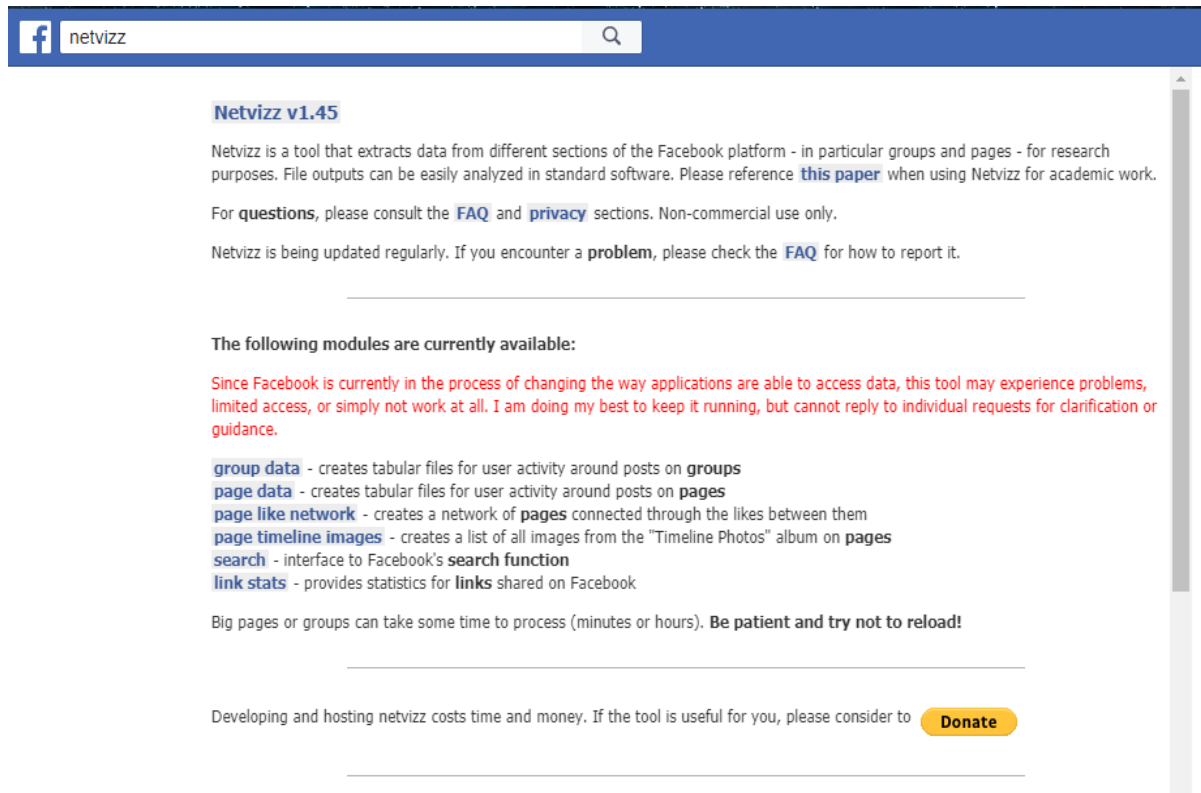


Figura 11. Tela inicial da ferramenta de coleta NETVIZZ.
Fonte: Próprio autor (2018).

Com o módulo *Page data* selecionado, se deve informar o *ID do Facebook* da página em questão, a quantidade de postagem que o aplicativo buscará ou o período e os dados que deseja se obter. Neste caso é possível selecionar três tipos de formas de extração, sendo elas *Post statistics only* (somente estatísticas de postagens), *Post statistics and 200 tops ranked comments per post* (estatísticas de postagem e os 200 melhores comentários por postagem) e *Full data* (dados completos de rede e comentários). É possível especificar se as postagens recuperadas serão apenas feitas pelo próprio pagina ou pelas páginas e usuários.

A figura 12 ilustra a inserção dos valores na ferramenta, que disponibiliza um site para encontra o ID do *Facebook* da página que se quer fazer o download dos dados bastando clicar na instrução *find page ids here*, e colocar a *url* da página em questão.

page id: (find page ids [here](#) or through Netvizz' [search module](#))

date scope:
 last posts (max. 999)

 posts between and

data to get:
 post statistics only (post metrics, stats per day and fans per country)

 post statistics and 200 top ranked comments per post

 full data (full network and comment files, can fail for larger pages)

get [post by page only](#) or [posts by page and users](#)

Figura 12. Tela de busca por dados de página no NETVIZZ.
Fonte: Próprio autor (2018).

Nesta pesquisa definiu-se o período entre 20 de dezembro de 2017 a 10 de abril de 2018 em virtude do período de férias, carnaval e semana santa onde o fluxo de turistas geralmente aumenta na região. A forma de extração dos dados foi definida como *Post statistics and 200 tops ranked comments per post*, por já trazer uma seleção das melhores postagens. E os dados capturados foram *posts by page and users* (tanto de postagem da página feitas pelo administrador quanto dos próprios usuários).

Após a definição dos dados que se quer obter, a aplicação carregará os dados e os disponibilizará para *download* em no botão *zip archive*, no formato zip (formato de compactação de arquivos). A Figura 13 demonstra a saída dos dados baixados e descompactados.

Nome	Tamanho	Comprimido	Tipo	Modificado	CRC32
Disco Local					
page_297111820335738_2018_07_12_22_09_54.tab	97.646	19.136	Arquivo TAB	12/07/2018 22:09	4E2B078E
page_297111820335738_2018_07_12_22_09_54_statsperday.tab	2.496	617	Arquivo TAB	12/07/2018 22:09	510BD156
page_297111820335738_2018_07_12_22_09_54_topcomments.tab	119.287	10.564	Arquivo TAB	12/07/2018 22:10	593AD4DE

Figura 13. Saída de dados após a descompactação.
Fonte: Próprio autor (2018).

Os arquivos descompactados contendo os dados referentes as páginas estão em formato de tabela e podem ser abertos por qualquer programa de leitura de planilhas eletrônicas como

o Microsoft *Exel* e o Google *Sheets*. O arquivo **.tab** contém uma tabela com todas as postagens feitas pela página no período pesquisado, os tipos de postagens, números de curtidas, números de comentários, reações, engajamento, link da publicação entre outros compôs. O arquivo **startspersday.tab** corresponde as estatísticas das diárias das postagens feitas, com os campos da data da publicação, número de curtidas, compartilhamentos, reações e o número de publicações feitas. O terceiro arquivo **topcomments.tab** diz respeito aos comentários dos usuários do *Facebook* que publicaram nas postagens feitas pela página.

4.2 ETAPA 2 - PRÉ-PROCESSAMENTO

Segundo Camilo e Silva (2009) é costumeiro que os dados não se apresentem prontos para aplicação de técnicas de mineração de dados diretamente. Para poder aplicar tais técnicas muitas vezes é necessário antes realizar um processo de limpeza dos dados, como aplicar filtros, combinações ou preencher os espaços vários, dependendo da qualidade dos dados que serão utilizados.

No caso deste estudo, pela natureza dos dados adquiridos, a etapa de limpeza e adequação dos dados para as etapas posteriores de mineração foi feita manualmente, removendo os comentários ilegíveis, marcações de outros usuários e espaços vazios. Além disso muitos comentários apresentavam caracteres inválidos, neste caso foi feita a correção destes caracteres de acordo com a língua portuguesa.

O Quadro 7 exemplificam as classificações adotadas na etapa de limpeza dos dados.

COMMENT_ID	COMMENT_MESSAGE	PUBLISHED
2073449359349801_2073497126011691	Eu amo! Muito lindo!	2018-04-06
2073449359349801_2075561629138574	Perfeito já fui.	2018-04-08
2073449359349801_2073674229327314	Deslumbrante! Amei	2018-04-07
2073449359349801_2073460289348708	Glória Maria	2018-04-06
2073449359349801_2074040455957358	Ricardo Amorim	2018-04-07
2072094946151909_2075620692466001	Eu também quero conhecer	2018-04-08
2036754119685992_2041358379225566	δÿ~□δÿ~□δÿ~□δÿ~□	2018-03-10
2032022470159157_2032070626821008		2018-03-02
2005355372825867_2033896019971802	Na rota das emoções passa pela praia da Tatajuba?	2018-03-04
2037734146254656_2071349382893132	Um passeio, mas lindos que eu já fiz vale muito a pena	2018-04-05

Quadro 7. Etapa de limpeza dos comentários.

Fonte: Próprio autor (2018).

Os comentários preenchidos com verde são considerados válidos para a análise de texto e geração da nuvem de palavras, os com preenchimento azul significam as marcações de outros usuários do *Facebook* e consequentemente os marcados em amarelo representa os comentários inválidos (campos vazios ou com caracteres ilegíveis).

4.3 ETAPA 3 – PROCESSAMENTO

O processamento de dados pode ser definido como o tratamento sistêmico de estruturas de dados, utilizando sistemas informatizados visando a ordenação e classificação entre outro tipo de tratamento que pode ser realizado sobre um dado ou conjunto de dados. Através de um plano programado anteriormente, seu objetivo é adquirir um determinado resultado que agregue conhecimento sobre os dados estudados.

Nesta etapa o processamento se dividiu em dois: mineração de texto e mineração de dados. Os dados textuais recuperados foram analisados seguindo as técnicas e procedimentos referentes à análise de sentimentos em texto. Os dados estatísticos foram transformados em nominais para análise de dados na ferramenta *WEKA*, com o objetivo descobrir as regras de associação referentes às postagens.

O processo de mineração de texto, utilizando a biblioteca *textblob* escrita python para análise de sentimentos, foi escrito um script em python, para analisar o sentimento nos comentários dos usuários nas postagens das páginas. O script mede os graus de polaridade e subjetividade de um comentário, diferenciando um comentário positivo, negativo ou neutro.

Os comentários são passados através de um arquivo de texto para o script que abre e faz a leitura do arquivo. As frases são passadas por meio de uma lista para a *textblob*, caso o idioma do comentário esteja em inglês ele passa direto para ser analisado sem a necessidade de conversão, caso contrário o comentário é traduzido para o inglês e posteriormente é analisado. Se a polaridade for maior de zero, então o comentário é avaliado como positivo, se o for menor que zero, o comentário é avaliado como negativo e caso seja igual a zero então o comentário é avaliado como neutro

A Figura 14 mostra o script que foi desenvolvido que roda na versão 3.6 do python foi a 3.6 utilizando o Ambiente de Desenvolvimento Integrado (IDE, do inglês *Integrated Development Environment*) PyCham.

```

1  #-*- coding: utf-8 -*-
2
3  import textblob # Faz a importação da Biblioteca Textblob
4  from textblob import TextBlob
5
6  text = open('textopt.txt','r') # Abri um arquivo .txt
7  cont = 0 # Iniciando variavel que servira para conta o numero da linha
8
9  # Loop para iterar nas linhas dos comentários
10 for i in text:
11     frase = TextBlob(i) # Frase recebe a linha e coloca no Textblob
12
13     #Caso idioma diferente de Inglês
14     # Faz a detecção do idioma do texto e faz a tradução automatica para o Inglês
15
16     if frase.detect_language() != 'en':
17         traducao = textblob.TextBlob(str(frase.translate(to='en')))
18         cont = cont + 1
19
20         if (traducao.sentiment.polarity > 0.1): #Para polaridade maior que 0.1 imprimi sentimento POSITIVO
21             print(cont,traducao.sentiment.polarity,traducao.sentiment.subjectivity, 'positive')
22
23         elif (traducao.sentiment.polarity == 0.0): #Para polaridade igual a 0.0 imprimi sentimento NEUTRO
24             print(cont,traducao.sentiment.polarity,traducao.sentiment.subjectivity, 'neutral')
25
26         else:
27             #Para polaridade menor que 0.0 imprimi sentimento NEGATIVO
28             print(cont,traducao.sentiment.polarity,traducao.sentiment.subjectivity, 'negative')
29
30     # Caso idioma esteja em Inglês
31     else:
32         cont = cont + 1
33         if (traducao.sentiment.polarity > 0.1): #Para polaridade maior que 0.1 imprimi sentimento POSITIVO
34             print(cont,traducao.sentiment.polarity,traducao.sentiment.subjectivity, 'positive')
35
36         elif (traducao.sentiment.polarity == 0.0): #Para polaridade igual a 0.0 imprimi sentimento NEUTRO
37             print(cont,traducao.sentiment.polarity,traducao.sentiment.subjectivity, 'neutral')
38
39         else:
40             #Para polaridade menor que 0.0 imprimi sentimento NEGATIVO
41             print(cont,traducao.sentiment.polarity,traducao.sentiment.subjectivity, 'negative')

```

Figura 14. Script em python para análise de polaridade e subjetividade em postagens de comentários no Facebook.

Fonte: Próprio autor (2018).

A subjetividade do texto também é um parâmetro retornado pelo script, neste caso os valores da subjetividade variam entre zero e um, indicando que quando o valor for um a subjetividade do texto será de 100%, e assim para os demais valores. Outro ponto interessante sobre o grau de subjetividade que foi observado durante os testes, que quanto mais próximo de 100% for a subjetividade de um texto, mais difícil para o script classificar os comentários, levando mais tempo para execução.

O processo de mineração de dados, utilizou para esta tarefa a ferramenta WEKA, por ser uma aplicação amplamente utilizada em trabalhos acadêmicos, principalmente pela ampla gama de algoritmos de descoberta de conhecimento oferecidos por ela. A técnica escolhida para ser utilizada no presente estudo, foi a associação utilizando o algoritmo *apriori*, aterrando apenas os valores de que definem o suporte mínimo *lowerBoundMinSupport* para 0.4, garantido assim que todas as regras geradas estejam em pelo menos 40% das transações.

A Figura 15 ilustra a tela da ferramenta após a execução da tarefa de mineração de regras de associação com o algoritmo *apriori*.

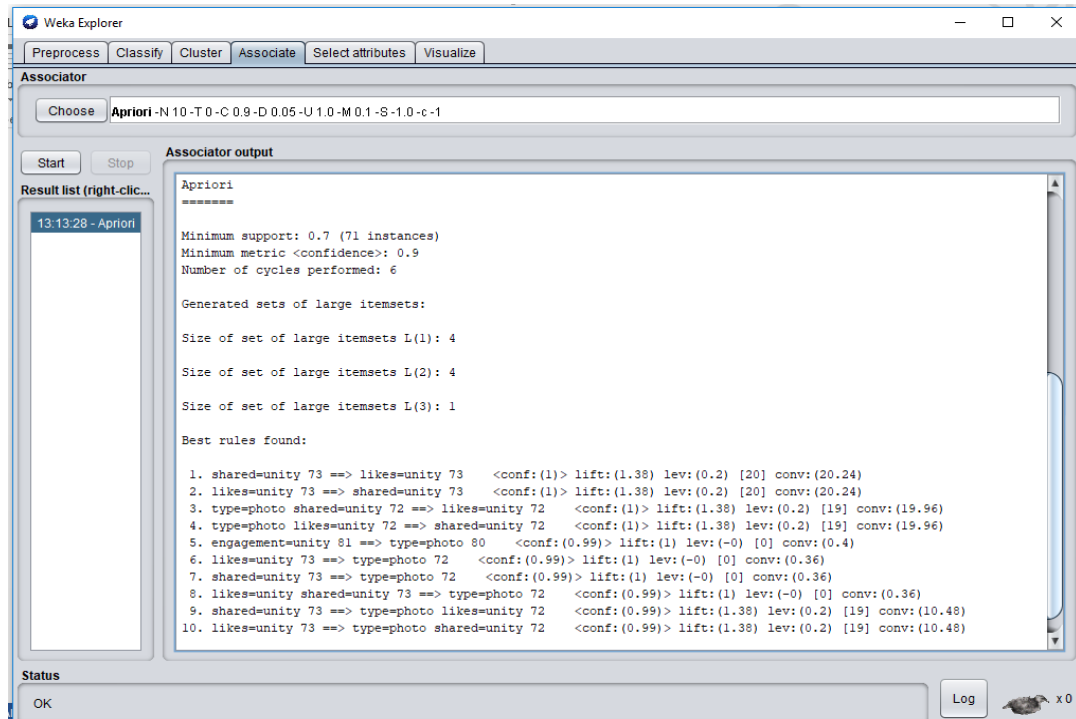


Figura 15. Execução do algoritmo a priori na ferramenta *Weka*.
Fonte: Próprio autor (2018).

O Quadro 8 apresenta os atributos, o significado e os valores respectivos, estes dados serão usados para a leitura das regras geradas no final do processo.

Atributos	Significado	Valor
Type	Tipo	Photo, Link, Status e Vídeo
Likes	Curtidas	Numérico
Comments	Comentários	Numérico
Shared	Compartilhado	Numérico
Engagement	Engajamento	Numérico
Love	Amor	Numérico
Haha	Risos	Numérico
Wow	Uau	Numérico
Sad	Triste	Numérico
Angry	Zangado	Numérico
Thankful	Agradecido	Numérico

Quadro 8. Atributos, significados e valores para a leitura das regras de associação geradas.
Fonte: Próprio autor (2018).

Os valores numéricos presentes nas postagens foram convertidos para nominais, para que fosse possível gerar as regras de associação, o Quadro 9 ilustra os valores numéricos e os respectivos valores nominais após a conversão feita para a geração das regras.

Numeral	Nominal	Significado
0	?	Vazio
1 a 9	Unity	Poucos
10 a 99	Dicke	Dezenas
100 a 999	Hundred	Centenas
1000 ou superior	Thousand	Milhares

Quadro 9. Conversão dos dados numéricos em nominais.
Fonte: Próprio autor (2018).

A Figura 16 ilustra os *emoticon*, presentes no *Facebook*, eles indicam também a reação dos usuários em relação a publicação.



Figura 16. Emoticons presentes no *Facebook*.
Fonte: Próprio autor (2018).

4.4 ETAPA 4 – ANÁLISE E VISUALIZAÇÃO OS RESULTADOS

Segundo Freitas (2008) a aplicação de técnicas de visualização de informação na análise de redes sociais, permitem realizar uma observação mais completa e interativa favorecendo os processos de descoberta de conhecimento. A exploração de tais técnicas permite visualizar o conhecimento extraído nas etapas anteriores de processamento, de maneira clara e intuitivas, facilitando a legibilidade e abstração da informação.

Nesta etapa, além da análise e visualização dos resultados encontrados pela mineração de sentimento em texto e mineração de regras de associação de estáticos das *fan pages* estudadas, também foi elaborado a visualização da rede de relacionamentos das agências analisadas. Utilizando a ferramenta *Gephi* na versão 0.9.2, um fato importante de se destacar, é necessário ter o Java instalado na máquina para utilizar a aplicação. Com ela foi possível visualizar as informações contidas no grafo das agências, identificar grupos (cluster) e evidenciar os elementos com os maiores números de graus.

Para gerar um grafo que possa ser interpretado, no menu estatísticas executa-se a função para calcular a modularidade, dividindo a rede em partes. No menu de partição aplica-se a divisão por modularidade para executar a separação dos módulos da rede, com este procedimento a rede se divide em cores diferente para cada módulo, mais a dispersão física se

mantém. A próxima etapa foi separar os nós da rede fisicamente, para esta tarefa existem diversos algoritmos no *Gephi*, porém foi usado neste exemplo o algoritmo *Force Atlas 2*, com o parâmetro evitar sobreposição devidamente marcado (VAFG, 2015).

A Figura 17 mostra a tela inicial do *Gephi* com uma *fur ball* (bola de pelos), ou seja, uma rede que ainda não foi trabalhada pelos algoritmos do *software*.

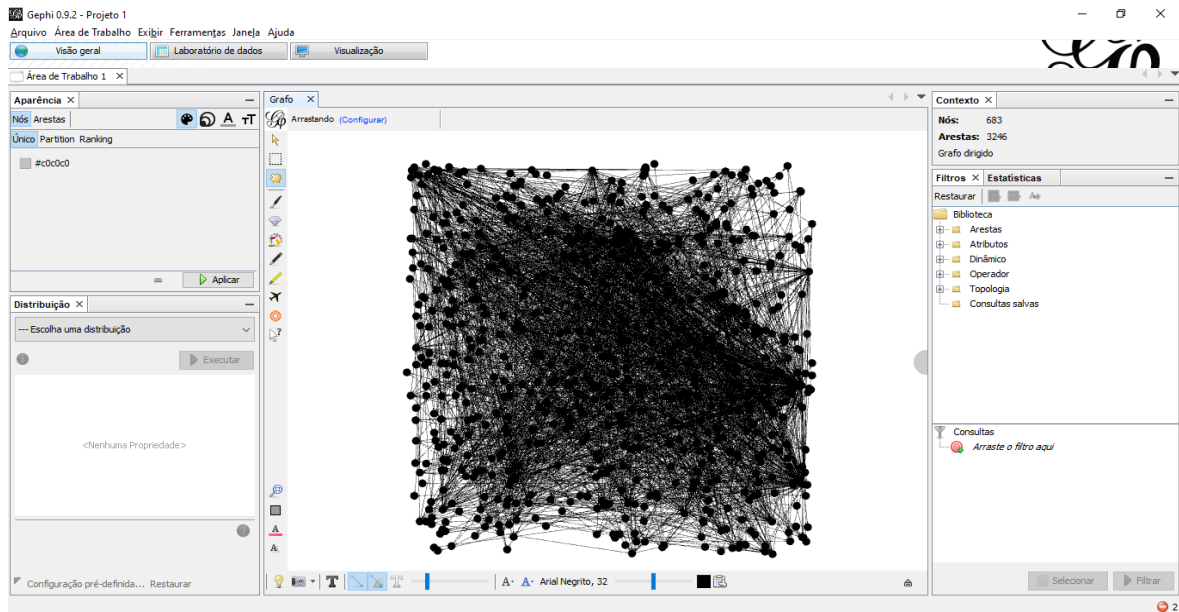


Figura 17. Visualização de uma rede antes de ser tratada na ferramenta gephi.
Fonte: Próprio autor (2018).

Após o tratamento feito, tornou-se possível visualizar e separar os *clusters*, além de evidenciar os principais nós das redes e as ligações entre os nós que são representados pelas arestas.

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Neste capítulo são apresentados os resultados alcançados aplicando a metodologia desenvolvida com a finalidade de extrair e analisar dados provenientes das *fan pages* do *Facebook*, aplicando as etapas supracitadas para descoberta de conhecimento em páginas de agências de turismo presentes no litoral do estado do Piauí. Assim como nos modelos desenvolvidos por Neves (2013) E Thomaz (2014) que aplicaram seus modelos para descobrir novos conhecimentos no setor do turismo, este trabalho se voltou para o setor turístico como forma de avaliar a metodologia proposta, muito embora os métodos expostos aqui, possam ser aplicados em qualquer outro tipo de página ou mesmo em grupos presentes na rede social em questão.

5.1 COLETA DE DADOS

Os dados coletados foram extraídos entre dezembro de 2017 e abril de 2018, utilizando a ferramenta Netvizz, para adquirir os dados das agências selecionadas. No total foram recuperadas 379 postagens de diferentes tipos feitas pelos administradores *fan pages*, e 1811 comentários feitos por usuários nas *fan pages* das agências, incluindo marcações de amigos feitas pelos usuários nas páginas e comentários vazios ou considerados inválidos para análise por apresentar caracteres indistinguíveis.

A Figura 18 mostra a porcentagem de cada *fan pages* nas postagens recuperadas.

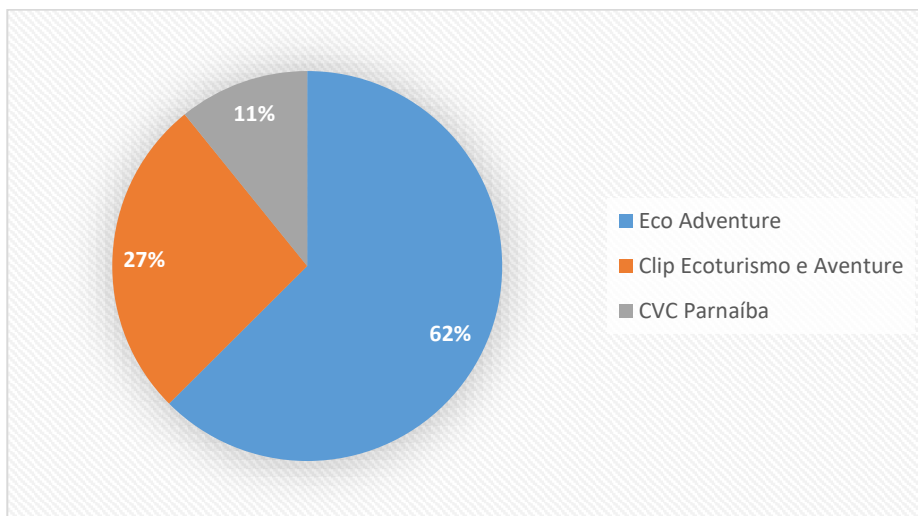


Figura 18. Porcentagem de cada *fan pages* do total de postagens recuperadas.

Fonte: Próprio autor (2018).

Deste total a Eco Adventure aparece em primeiro com 237 postagens do total de postagens recuperadas representando 62% do total, a Clip Ecoturismo e Aventure em segundo com 101 postagens representando 27% do tal e a CVC Parnaíba com 41 postagens representando 11% do número de postagens recuperadas pela ferramenta. Estas postagens forneceram os dados estatísticos que serviram de entrada para a mineração de regras de associação que utilizou a ferramenta *WEKA* para realizar as análises.

Quanto aos dados referentes aos comentários dos usuários, como já foi exposto foram recuperados 1811 comentários no igual período. A Figura 19 apresenta os dados referentes aos comentários recuperados como percentual de comentários analisados, usuários que marcaram amigos nas postagens e o percentual de comentários inválidos.

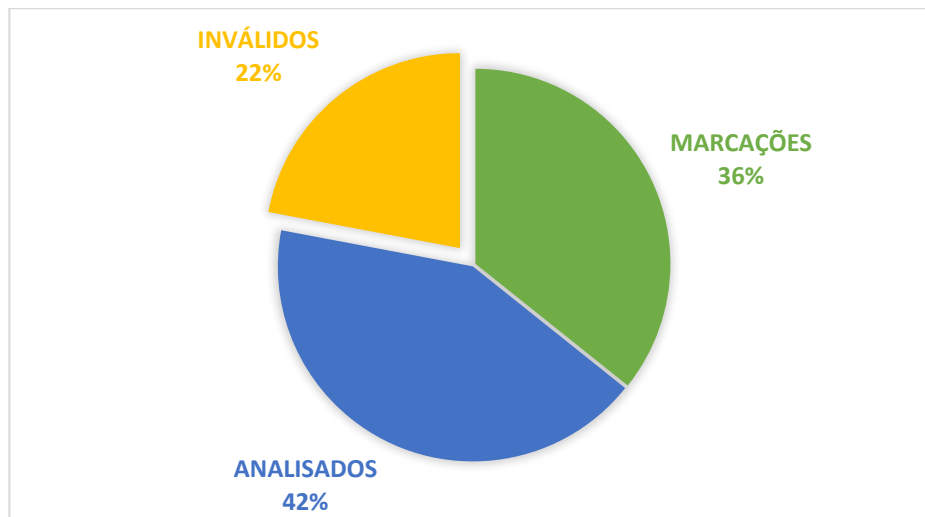


Figura 19. Percentual de comentários analisados.
Fonte: Próprio autor (2018).

Foram analisados 765 comentários, representando 42% do percentual total, não foram incluídos na análise as marcações de amigos nos comentários, pois as não apresentam textos que possam ser passados para a análise de sentimentos textos, ainda sim elas representam um fator importante pois elas atuam como uma maneira de divulgação das postagens das *fan pages* e conseqüentemente dos produtos e serviços oferecidos, estas do total de comentários recuperados representaram 36% do total, sendo um número bastante significativo para o marketing das organizações. Os comentários vazios ou inválidos foram 22% do total, mesmo sendo um percentual um pouco alto, não houve o comprometimento das etapas de coleta nem das demais etapas que se utilizaram dos dados obtidos.

5.2 RESULTADOS DOS PROCESSOS DE MINERAÇÃO

Após a limpeza e adequação dos dados o primeiro processamento realizado em cima do conteúdo recuperado foi a análise de sentimentos em texto. Nesta etapa buscou-se distinguir a opinião dos usuários sobre o conteúdo postado nas *fan pages*. Utilizando o script desenvolvido em python para este fim, descrito na seção 4.3, do total de 1811 comentários recuperados, 42% foram analisados totalizando 764 postagens submetidas ou script.

O Quadro 10 mostra a quantidade de postagens analisadas de cada *fan pages* estudada, além das marcações de amigos feitas por usuários nas páginas e o número de comentários inválidos.

	MARCAÇÕES	ANALISADOS	INVÁLIDOS	TOTAL DE COMENTÁRIOS
Clip Ecoturismo	5	129	1	135
CVC Parnaíba	416	317	314	1047
Eco Adventure	227	318	84	629
TOTAL	648	764	399	1811

Quadro 10. Números gerais de comentários analisados, marcações e postagens inválidas.

Fonte: Próprio autor (2018).

Os dados foram analisados e classificados pelo script de acordo com o grau de polaridade e subjetividade em negativo, neutro e positivo. De modo geral os resultados observados apontam para uma visão positiva das páginas pelos usuários, onde apresentam pequenos números de comentários negativos. Os comentários que foram classificados como neutros, em sua maioria correspondem a dúvidas dos sobre locais e horários de funcionamento, roteiros, respostas a outros comentários e perguntas sobre valores. Os comentários positivos correspondem a elogios a equipe da empresa e aos roteiros.

O Quadro 11 exemplifica os valores retornados pela polaridade, subjetividade, avaliação do comentário e os comentários avaliados pelo script.

Nº	Polaridade	Subjetividade	Avaliação	Comentários
1	0.90	0.78	Positivo	Muito bom esse passeio.
2	0.80	0.75	Positivo	Ótimo passeio, recomendo.
3	0.71	0.73	Positivo	Muito bom e bonito este meu Piauí, muitos lugares bonitos.
4	0.0	0.50	Neutro	É aberto durante a semana?
5	0.0	0.0	Neutro	Oi, onde fica esta marina?
6	0.70	0.77	Positivo	Gostei muito do local, ótimo atendimento e vista privilegiada.

7	-0.23	1.0	Negativo	Não fui porque não disseram onde fica “aff”.
8	0.0	0.0	Neutro	Onde fica?
9	1.0	1.0	Positivo	Muito lindo.
10	0.23	0.77	Positivo	Meu Piauí, que, com belíssimas paisagens!

Quadro 11. Valores retornados pelo script de análise de sentimentos em textos.

Fonte: Próprio autor (2018).

As 764 postagens analisadas possibilitaram a observação da visão dos usuários sobre os serviços oferecidos pelas 3 agências analisada. Sem assim a Figura 20 demonstra os resultados da análise de sentimentos sobre os comentários feitos pelos usuários nas páginas.

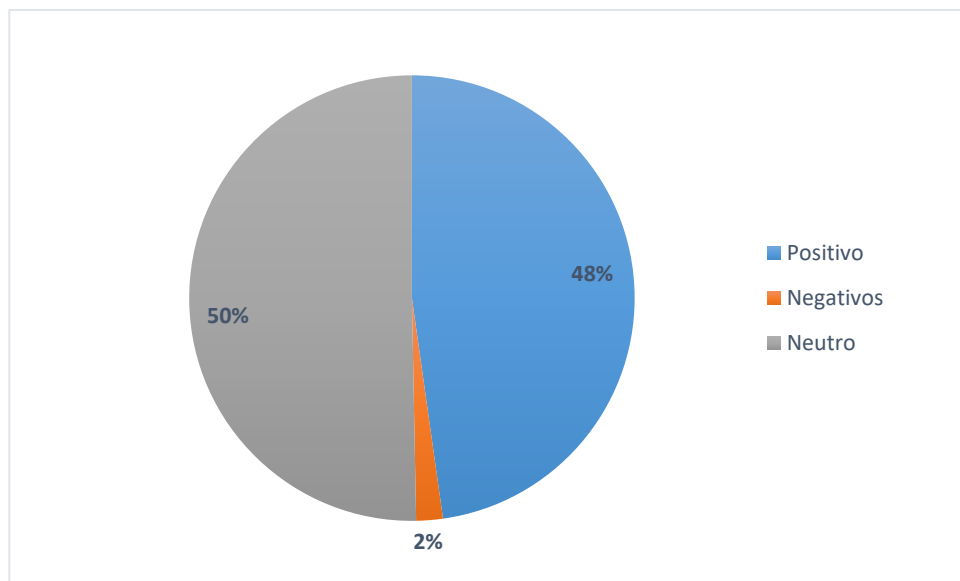


Figura 20. Sentimento percebido nos comentários pela análise de sentimentos em textos.

Fonte: Próprio autor (2018).

Os resultados obtidos pós-processamentos mostraram que a maioria dos comentários foram classificados como neutros, representando 50% (417 comentários) de todos os comentários avaliados, com uma pequena margem acima dos comentários classificados como positivos que alcançaram 48% (396 comentários), os negativos representaram apenas 2% (16 comentários). Muito embora esperava-se a que a maioria dos comentários fossem classificados com positivos, este resultado demonstra a boa aceitação e qualidade nos serviços prestados pelas agências, uma vez que os comentários classificados como neutros, muitas vezes expressaram o interesse pelos produtos ou serviços oferecidos pelas organizações, porém não foram classificados como positivos por não expressar a satisfação ou contentamento pelos serviços, mais sim revelando o interesse que poderia ser convertido ou não na aquisição do serviço por parte do usuários.

Por fim a Figura 21 apresenta a quantidade individual de comentários negativos, positivos e neutro classificados em cada agência.

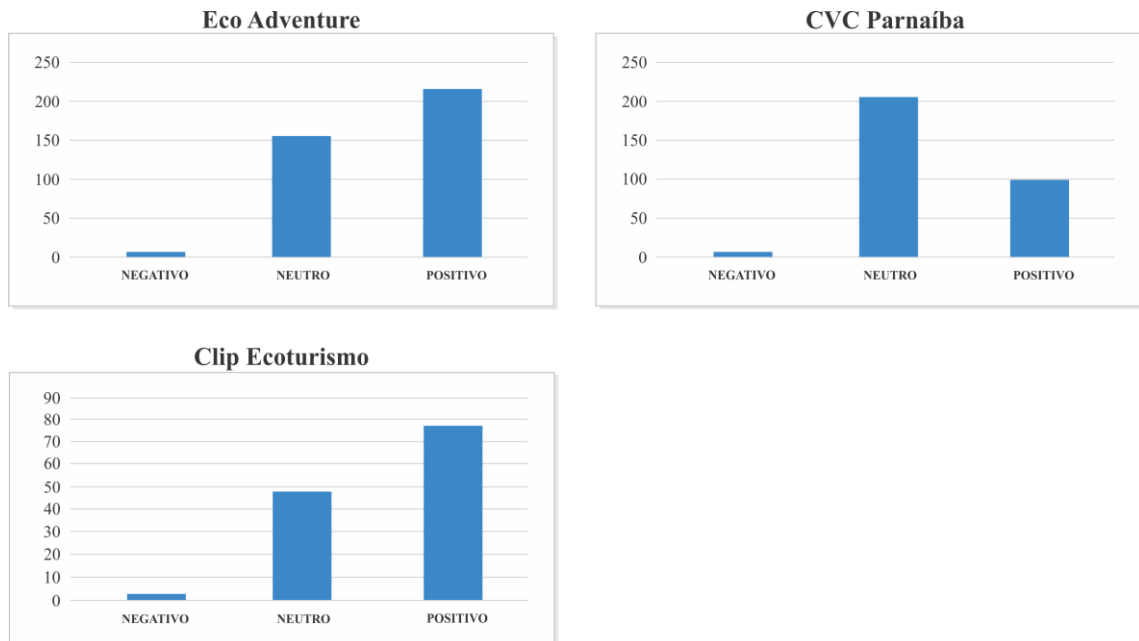


Figura 21. Análise de sentimentos das agências presentes no litoral do piauiense.
Fonte: Próprio autor (2018).

A organização que apresentou o maior número de comentários positivos foi a Eco Adventure com 220. Em relação aos comentários neutros a CVC Parnaíba obteve o maior número de comentários classificados com 212 e o maior número de comentários negativos foi observado na Eco Adventure com 9, também a que teve a maior quantidade de comentários analisados com 318.

O segundo tipo de processamento feito em sobre os dados extraídos das *fan pages*, foi a mineração de regras de associação. Embora a quantidade de dados analisados tenha sido pequena para esta técnica de mineração de dados, pois geralmente os padrões mais relevantes são mais facilmente percebidos quando tratamos de grandes bases de dados, os padrões observados podem ser trabalhados e analisados de forma a agregar valor aos demais processos realizados, além disso, aplicando a técnica em páginas maiores consequentemente o volume de dados irá aumentar, melhorando os resultados obtidos pela análise.

Com os dados estatísticos advindos das *fan pages* foi realizado a mineração de regras de associação, onde a preparação dos dados envolveu a adaptação para aplicação adequada da técnica tendo em vista que as regra de associação trabalham com dados nominais, e os dados que serviram com parâmetros de entrada são dados numéricos.

Foram geradas 100 regras para cada página, e selecionadas as 10 melhores regras para que se pudesse entender o como os meios mais eficazes para impulsionar o crescimento das *fan pages*.

Os Quadros 12 mostram as 10 principais regras associação encontradas na página da Clip ecoturismo, com o algoritmo *apriori*.

Clip Ecoturismo	
Associação	Confiança (%)
SE o tipo foto e compartilhado pouco ENTÃO curtida pouco	100
SE dezena de curtida ENTÃO tipo foto	100
SE tipo foto e compartilhado dezena ENTÃO curtida dezena	100
SE compartilhado dezena E amor dezena ENTÃO tipo foto	100
SE curtida dezena E compartilhado dezena E engajamento pouco ENTÃO tipo foto	100
SE compartilhado dezena E engajamento pouco E amor dezena ENTÃO curtida dezena	100
SE comentário pouco E amor dezena ENTÃO engajamento pouco	98
SE amor pouco ENTÃO curtida pouca	98
SE amor pouco ENTÃO compartilhado pouco	100
SE curtida dezena E compartilhado dezenas E engajamento pouco ENTÃO amor dezena	98

Quadro 12. Top 10 regras de associação na página da Clip Ecoturismo.

Fonte: Próprio autor (2018).

O Quadro 13 apresenta as 10 principais regras de associação geradas a partir da análise das postagens feitas pela *fan pages* da CVC Parnaíba.

CVC Parnaíba	
Associação	Confiança (%)
SE curtida centena ENTÃO TIPO FOTO	100
SE amor centena ENTÃO curtida centena	100
SE tipo foto E curtida centena E amor centena ENTÃO compartilhada centena	100
SE risos dezena ENTÃO curtida centena	100
SE risos dezena ENTÃO compartilhada centena	100
SE curtida centena E compartilhada centena E risos dezena ENTÃO tipo foto	100
SE comentado dezena E compartilhado centena ENTÃO amor centena	100
SE comentado dezena E compartilhado centena ENTÃO tipo foto	100
SE tipo foto E amor centena E risos dezena ENTÃO compartilhada centena	100
SE tipo foto E curtida centena E comentado dezena ENTÃO amor centena	100

Quadro 13. Top 10 melhores regras encontradas na página da CVC Parnaíba.

Fonte: Próprio autor (2018).

O Quadro 12 expõe as principais regras de associação geradas pela a ferramenta *WEKA* utilizando como entrada os dados recuperados nas páginas a Eco Adventure.

Eco Adventure	
Associação	Confiança (%)
SE curtida dezena E amor dezena ENTÃO compartilhada dezena	100
SE curtida pouco E amor pouco ENTÃO compartilhada pouco	98
SE tipo link E compartilhada dezena ENTÃO amor dezena	98
SE tipo link E curtida dezena E engajamento pouco ENTÃO compartilhada dezena E amor dezena	100
SE tipo foto E curtida dezena ENTÃO compartilhada dezena	93
SE curtida centena ENTÃO amor centena	98
SE comentada pouco E compartilhada pouco E engajamento pouco ENTÃO curtida dezena	100
SE engajamento pouco E amor dezena E triste pouco ENTÃO compartilhada dezena	98
SE compartilhada dezena E risos pouco ENTÃO amor dezena	93
SE comentada pouco E engajamento pouco E amor dezena E risos pouco ENTÃO compartilhada dezena	93

Quadro 14. Top 10 regras de associação geradas a partir da página da Eco Adventure.

Fonte: Próprio autor (2018).

Apesar da pequena quantidade de regras geradas e dos poucos dados disponíveis para gerá-las, é possível observa o valor do conhecimento que a aplicação desta técnica pode trazer para as organizações. Tais conhecimentos são capazes de direcionar para as melhores tomadas de decisões dentro das empresas, no que diz respeito ao marketing digital praticado por elas nas suas *fan pages*, impulsionando suas publicações aumentando a visibilidade e atraindo a atenção para os seus serviços e produtos.

Um exemplo do conhecimento gerado pelas regras apresentadas nos Quadros, podemos citar: se uma postagem tiver pouco compartilhamentos ela conseqüentemente terá poucas curtidas; se uma postagem for do tipo foto e obtiver poucos comentários e poucas reações de amor pela postagem então o engajamento naquela postagem também seja pouco; se uma postagem for do tipo foto, tiver centenas de reações de amor pela postagem e desenhadas de reações de risos nas postagens então ela será compartilhada centenas de vezes. Estes são só alguns exemplos do conhecimento extraídos das associações citadas que pode ser utilizado para impulsionar a visibilidade dentro das redes sociais.

5.3 VISUALIZAÇÃO DA REDE DE RELACIONAMENTO

A análise visual desempenha um papel crucial para entendermos melhor as *fan pages*, uma vez que a estrutura fundamental de uma rede social, se assemelha muito com a estrutura dos grafos, com seus nós e arestas. Esta fase se dividiu em duas partes, sendo estas: visualização das nuvens de palavras com os comentários dos usuários; e visualização do grafo de relacionamento das *fan pages*.

A geração das nuvens de palavras com os comentários dos usuários foi realizada pela ferramenta *web* Tagul, que possibilita a criação de estruturas que evidenciam os termos que aparecem com maior frequência. A Figura 22 ilustra a nuvem de palavras gerada pela ferramenta, para os comentários dos usuários na página da Clip Ecoturismo, onde as palavras que merecem destaque são: “Lindo”, “Paraíso”, “Delta”, “Piauí” e “conhecer”.



Figura 22. Nuvem de palavras dos comentários da Clip Turismo.

Fonte: Próprio autor (2018).

A Figura 23 apresenta os termos mais presente nos comentários dos usuários na página da CVC Parnaíba, destacando a presença das palavras “Lindo”, “Quero”, “Pacote”, “vamos”, “Casal” e “Valor”.



Figura 23. Nuvem de palavra mais frequentes nos comentários dos usuários na página da CVC Parnaíba.
Fonte: Próprio autor (2018).

A Figura 24 ilustra a nuvem gerada para os comentários na página da Eco Adventure, destacando a presença das palavras “Lindo”, “Amei”, “Saudade”, “Muito” e “Conhecer”.



Figura 24. Nuvem de palavras geradas pelos comentários na fan pages da Eco Adventure.
Fonte: Próprio autor (2018).

A visualização do grafo de relacionamento das *fan pages*, possibilitou identificar os agrupamentos além das principais *fan pages* relacionadas com as investigadas. No caso, só foi possível gerar os resultados visuais das *fan pages* da Clip turismo e Eco Adventure pois a página da CVC Parnaíba não apresentou dados suficientes para geração de uma visualização da rede.

A Figura 25 mostra o grafo da Clip turismo tratado na ferramenta gephi. Foram identificados 7 *clustering* (Agrupamentos), distintos pelas cores, além disso a ferramenta contou a presença de 683 nós e 3246 arestas.

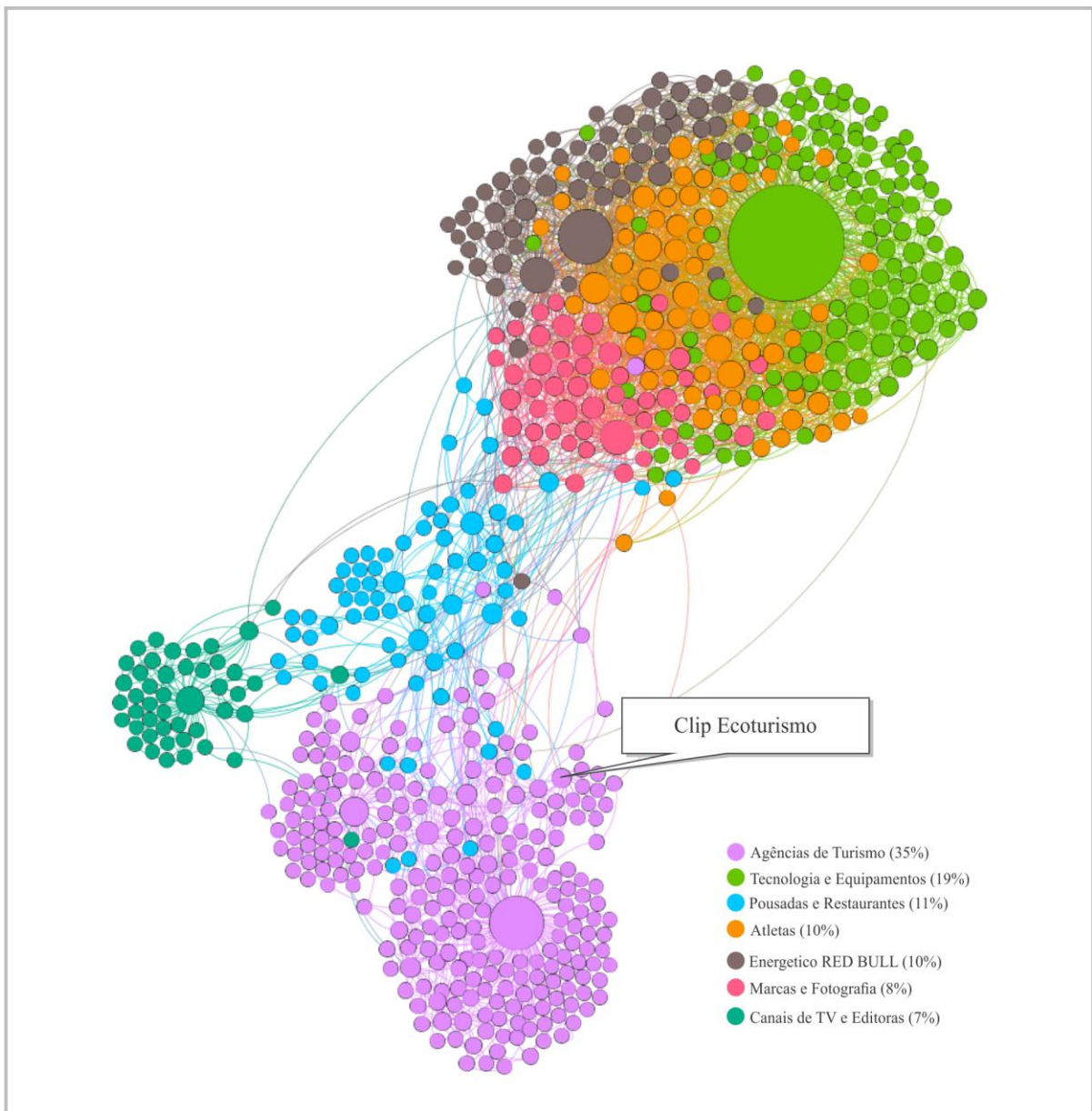


Figura 25. Rede de relacionamento da Clip Ecoturismo.

Fonte: Próprio autor (2018).

Foram identificados 7 *clustering* (Agrupamentos), distintos pelas cores, além disso a ferramenta contou a presença de 683 nós e 3246 arestas. Destaca-se a presença de inúmeras páginas de outras agências, sendo assim o maior cluster distinguido na cor lilás. Outro destaque é a *fan pages* como o maior grau entre todas as outras, a *fan pages* da GoPro identificada pela cor verde clara.

A Figura 26 ilustra a rede da Eco Adventure com suas páginas relacionadas.

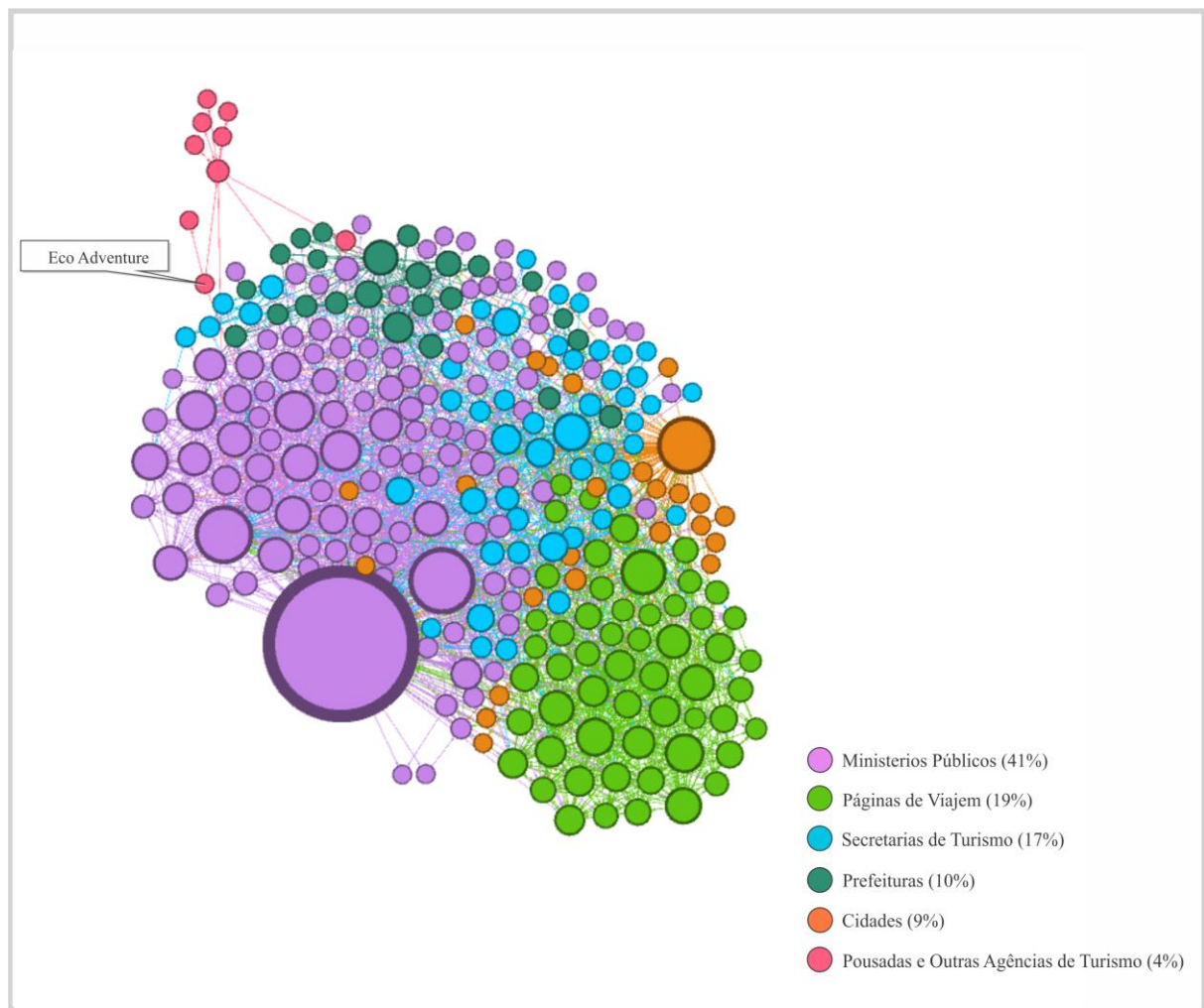


Figura 26. Rede de relacionamento da Eco Adventure.

Fonte: Próprio autor (2018).

Em destaque aparece a *fan pages* com o Maior grau entre todas as outras, a rede do ministério do turismo, evidenciada pelo tamanho e pela cor lilás com um anel na cor roxa na superfície. A rede como um todo contou com 279 nós e 2342 arestas e a presença de diversas páginas ligadas a prefeituras, secretarias e roteiros de viagens em diversas cidades.

6 CONSIDERAÇÕES FINAIS

Por meio da presente pesquisa, foi possível observar a relevância que as redes sociais têm para as organizações, em especial o *Facebook* por meio de suas *fan pages*, se tornando uma poderosa ferramenta para as organizações e consumidores trocarem informações e interagirem entre si. Os benefícios e oportunidades que a exploração dessa rede social oferece para as organizações vão além das práticas de marketing digital, atuando como canal de comunicação direta com o consumidor.

A aplicação da metodologia proposta por este trabalho possibilitou a análise por diferentes óticas sobre os dados extraídos das páginas. Isso porque, a adequação das técnicas aos dados extraídos viabilizou a análise mais abrangente e a utilização de plena dos dados extraídos.

Para verificar a eficiência do modelo, aplicou-se a metodologia proposta no setor de turismo, para assim averiguar o sentimento dos usuários do *Facebook*, que em algum momento publicaram comentários nas páginas das agências pesquisadas, além de identificar quais as reações dos usuários aparecem em conjunto mais frequentemente e a relação entre curtidas, compartilhamento e tipos de postagens mais usados pelas páginas analisadas.

Os resultados obtidos poderão ser melhorados aumentando a quantidade de dados extraídos, automatizando o processo de limpeza dos dados e cruzando os dados das postagens com os comentários publicados pelos usuários, acredita-se que com os cruzamentos dos dados estatísticos com a análise de sentimento possa-se gerar associações que relacione os números de curtidas e compartilhamentos por exemplo, com os sentimentos positivos ou negativos indicando quais postagens alcançarão mais visualizações.

Este estudo mostrou que as *fan pages* podem ser utilizadas como uma poderosa ferramenta para compreender melhor o sentimento do consumidor em relação às organizações, através da geração de conhecimento pelos dados que os próprios consumidores deixam nas páginas quando curtem, comentam ou compartilham seus conteúdos.

Como trabalhos futuros, pode-se estudar a criação de um *dataset* em português com palavras segmentadas ao contexto da *fan page* que se pretende estudar, a aplicação da técnica de sumarização de textos para criar um resumo com as principais palavras das postagens que possuem mais visualizações e a utilização de algoritmos de predição para projetar os números que um determinado tipo de postagem pode alcançar.

REFERÊNCIAS

- ABDEL-HAFEZ, A.; XU, Y. A Survey of User Modelling in **Social Media Websites**. v. 6, n. 4, p. 59–71, 2013.
- ABRAHAMS, A. S. et al. No TitleWhat's buzzing in the blizzard of buzz? Automotive component isolation in social media postings,. **Decision Support Systems**, v. 55, n. 4, p. 871–882, 2013.
- AGUIAR, A. **O que é, como criar e administrar uma Fanpage: o guia definitivo**. Disponível em: <<https://marketingdeconteudo.com/fanpage/>>. Acesso em: 8 de mai. de 2018.
- ALVES, M. M. **O papel das redes sociais no turismo**. p. 1–153, 2011.
- BARABÁSI, A.-L. **Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life**. 2003.
- BARCELLOS, M. A. R. **Avaliação da eficiência de comunicação em redes sociais digitais: uma análise exploratória do cenário de empresas no Brasil**. FUNDAÇÃO GETULIO VARGAS, 2010.
- BARION, E. C. N. MINERAÇÃO DE TEXTOS. **Revista de Ciências Exatas e Tecnologia**, v. III, p. 14-1-14–22, 2008.
- BASTIAN, M.; HEYMANN, S. **Gephi: An Open Source Software for Exploring and Manipulating Networks**. 2009.
- BATTAGLIA, M. DA G. B. **A Inteligência Competitiva modelando o Sistema de Informação de Clientes**. p. 200–214, 1999.
- BORGES, L. E. **Python para Desenvolvedores**. Rio de Janeiro, 2010.
- CABENA, P. et al. **Discovering Data Mining: From Concept to Implementation**. 1998.
- CAMILO, C.; SILVA, J. **Mineração de Dados: Conceitos, tarefas, métodos e ferramentas**. Universidade Federal de Goiás (UFG), p. 29, 2009.
- CASTELLS, M. **A Galáxia Internet: reflexões sobre a Internet, negócios e a sociedade**, 2009.
- CERVI, C. R. **Um Estudo sobre Mineração de Dados em Redes Sociais**. São Paulo: Atlas, 2008.
- CHAPMAN, P. et al. Step-by-step data mining guide. **CRISP-DM Consortium**, p. 76, 2000.
- CORRÊA, G. N. ET AL. **Uso da mineração de textos na análise exploratória de artigos científicos**. Disponível em: <http://www.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_3%0A83.pdf>. Acesso em: 5 jan. 2018.

CROOKS, A. et al. Earthquake: *Twitter* as a Distributed Sensor System. **Transactions in GIS**, v. 17, n. 1, p. 124–147, 2013.

CRUZ, V. L. L. DA et al. Redes sociais como estratégia de marketing turístico: o *Facebook* e a região norte do estado do Piauí-Brasil. **Turydes - revista de investigación en turismo y desarrollo local**, v. 5, n. 13, 2012.

DEGENNE, A.; FORSE, M. **Introducing Social Networks**. Sage Publications, 1999.

DENCKER, A. DE F. M. **Métodos e técnicas de pesquisa em Turismo**. Rio de Janeiro: Editora Futura, 2001.

DUQUE, C. B.; COSTA, D. V. F. **Gestão do conhecimento**: uma vantagem competitiva. Cleudina: II CONGRESSO DE ADMINISTRAÇÃO E TECNOLOGIA – CAT, p. 1–14, nov. 2016.

EMIRBAYER, M.; GOODWIN, J. Network Analysis , Culture , and the Problem of Agency. **American Journal of Sociology**, v. 99, n. 6, p. 1411–1454, 1994.

ERDÖS, P.; RÉNYI, A. On The Evolution Of Random Graphs. *Facebook*. **Graph API**. Disponível em: <<https://developers.facebook.com/docs/graph-api/>>. Acesso em: 23 de mai. 2018, 1960

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37, 1996.

FRANKENTHAL, R. **Comportamento do consumidor nas redes sociais**. Disponível em: <<https://mindminers.com/marketing/comportamento-do-consumidor-redes-sociais/>>. Acesso em: 23 de mai. 2018.

FREITAS, C. M. D. S. Extração de Conhecimento e Análise Visual de Redes Sociais. **Anais do xxviii Congresso da sbc**, p. 106–120, 2008.

GAMA, M. X. B. et al. A Liderança na Era da Informação e do Conhecimento nas empresas. **Research, Society and Development**, v. 3, n. 1, p. 2–18, 2016.

GRANOVETTER, M. **The Strength of Weak Ties: A Network Theory Revisited** Sociological Theory, 1983.

HALL, M. A. et al. The *WEKA* data mining software: an update. **SIGKDD Explorations**, v. 11, n. 1, p. 10–18, 2009.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. HAND, D. et al. **Principles of data mining**. 2008

HEA, W.; ZHA, S.; LI, L. Social media competitive analysis and text mining: A case study in the pizza industry. **International Journal of Information Management**, v. 33, n. 4, p. 464–472, 2013.

HEKIMA. **Como escolher um sistema de mineração de dados eficiente?** Disponível em:

<<http://www.bigdatabusiness.com.br/como-escolher-um-sistema-de-mineracao-de-dados-eficiente/>>. Acesso em: 23 de mai. 2018.

JANNUZZI, C. S. C.; FALSARELLA, O. M.; SUGAHARA, C. R. **Gestão do conhecimento**: um estudo de modelos e sua relação com a inovação nas organizações. p. 97–118, 2016.

KALAMPOKIS, EVANGELOS TAMBOURIS, EFTHIMIOS TARABANIS, K. Understanding the predictive power of social media. **Internet Research**, v. 23, n. 4, p. 544–559, 2013.

KAPLAN, ANDREAS M. HAENLEIN, M. Users of the world, unite! The challenges and opportunities of Social Media,. **Business Horizons**, v. 53, n. 1, p. 59–68, 2010.

KIETZMANN, J. H. et al. Social media? Get serious! Understanding the functional building blocks of social media. **Business Horizons**, v. 54, n. 3, p. 241–251, 2011.

LONGHI, C. **Análise de dados para apoio a tomada de decisão com ênfase no marketing digital**. Caxias do Sul: Aquarela, 2016.

LONGO, R. M. J.; QUEIROZ, C.; SANTOS, P.; CAMACHO, F.; PAULINELLI, R.; FEDELE, D. **Gestão do conhecimento: a mudança de paradigmas empresariais no século XXI**. São Paulo, 2014.

LUCHESE, F. E. S. **Gestão do conhecimento nas organizações - Companhia de engenharia de Tráfego**. São Paulo, 2012.

MANNARA, B. **Crie sua nuvem de palavras personalizadas com o Tagul**. Disponível em: <<http://www.techtudo.com.br/tudo-sobre/tagul.html>>. Acesso em: 2 jul. 2018.

MARTELETO, R. M. **Análise de redes sociais – aplicação nos estudos de transferência da informação**. 2001.

MILGRAM, S. **Behavioral Study of Obedience no Journal of Abnormal and Social Psychology**, 1963.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de Textos** GoiâniaUFG, , 2007.

NEVES, A. J. W. A.; MACHOTA, V. P. **Uso do Facebook como ferramenta de marketing nos meios de hospedagem do município de matinhos (PR)**. Foz do Iguaçu: Aquarela, 2016.

NEVES, A. J. W. A. DA. **Qualidade percebida de produtos e serviços turísticos em eventos**: uso de ferramentas para análise de conteúdo do *Twitter*. Curitiba: Universidade Federal do Paraná, 2013.

O'REILLY, T. **What Is Web 2.0**. Disponível em: <<https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>>. Acesso em: 25 mar. 2018.

PAL, S. K. .; TALWAR, V.; MITRA, P. Web mining in soft computing framework:

Relevance, state of the art and future directions. **Neural Networks, IEEE Transactions on**, v. 13, n. 5, p. 1163–1177, 2002.

PUREVIAGEM. **DICAS DE VIAGEM: LITORAL DO PIAUÍ, O MENOR LITORAL DO BRASIL**. Disponível em: <http://www.pureviagem.com.br/noticia/dicas-de-viagem-litoral-do-piaui-o-menor-litoral-do-brasil_a21605/1>. Acesso em: 1 jun. 2018.

PYTHON. **Linguagem Python**. Disponível em: <<https://www.python.org>>. Acesso em: 27 maio. 2018.

REZ, R. **A importância de uma fanpage no Facebook**. Disponível em: <<https://novaescolademarketing.com.br/comunicacao-corporativa/importancia-de-uma-fanpage-no-facebook/>>. Acesso em : 25 mai. 2018

REZENDE, S. O. Mineração de dados. **Dados**, v. 1, p. 397–433, 2003.

RIBEIRO, D. **Extraia gratuitamente dados de usuários, páginas e grupos do Facebook**. Disponível em: <<http://www.techtudo.com.br/tudo-sobre/netvizz.html>>. Acesso em: 1 jul. 2018.

RIEDER, B. Studying *Facebook* via data extraction: the Netvizz application. **DL.Acm.Org**, 2013.

SAFKO, L. **The Social media bible: tactics, tools, and strategies for business success**. 2010.

SANTOS, A. P. et al. **Data Mining em Redes Sociais**. Disponível em: <http://www.inf.ufsc.br/~luis.alvares/INE5644/G8_RedesSociais.pdf>. Acesso em: 25 mar. 2018.

SANTOS, D. C. DOS. **Coleta Automatizada E Análise De Dados Em Fan pages Do Facebook**. Curitiba. 2014

SANTOS, R. **Conceitos de Mineração de Dados na Web**. Disponível em: <<http://www.lac.inpe.br/~rafael.santos/Docs/WebMedia/2009/webmedia2009.pdf>>. Acesso em: 18 mar. 2018.

SANTOS, R.; PRÓSPERO, R. P.; GIANORDOLI, G. Estudo sobre a abordagem quantitativa para visualização de dados qualitativos: processo e ferramentas. **6º Information Design International Conference**, v. 1, p. 1–9, 2014.

SILVA, T. DO N. DA. **Uma arquitetura para descoberta de conhecimento a partir de bases textuais**. p. 71, 2012.

SILVA, F. B.; ROBERTO, P.; OLIVEIRA, M. DE. Práticas empreendedoras: uma análise das redes sociais como forma de impulsionar a gestão de vendas e os negócios das empresas. **Revista Conbrad**, p. 177–190, 2016.

SIQUEIRA, A. **Fanpage no Facebook: o que é e por que sua empresa deve ter uma**. Disponível em: <<https://resultadosdigitais.com.br/blog/fanpage-facebook/>>. Acesso em: 20 jun. 2018.

SONG, F. W. Theorizing web 2.0: A cultural perspective. **Information, Communication & Society**, v. 13, p. 249–275, 2010.

STATISTICS. **Global social networks ranked by number of users 2018**. Disponível em: <<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>>. Acesso em: 22 maio. 2018a.

STATISTICS. **Number of monthly active Facebook users worldwide as of 1st quarter 2018 (in millions)**. Disponível em: <<https://www.statista.com/statistics/264810/number-of-monthly-active-Facebook-users-worldwide/>>. Acesso em: 22 maio. 2018b.

TAVARES, W.; PAES, P. Movimentos Sociais em Redes Sociais Virtuais : Possibilidades de Organização de Ações Coletivas no Ciberespaço. p. 213–234, 2015.

TEXTBLOD. **TextBlod Análise de sentimentos**. Disponível em: <<https://textblob.readthedocs.io/en/dev/index.html>>. Acesso em: 27 maio. 2018.

THOMAZ, G. M. **Processo de mineração de conteúdos em mídias sociais para auxílio na gestão de destinos turísticos**. UNIVERSIDADE FEDERAL DO PARANÁ, 2014.

TORRES, C. **A Bíblia do Marketing Digital: Tudo o que você queria saber sobre o marketing e publicidade na internet e não tinha a quem perguntar**. São Paulo: Novatec, 2009.

TSURUOKA, Y.; TSUJII, J.; ANANIADOU, S. FACTA: A text search engine for finding associated biomedical concepts. **Bioinformatics**, v. 24, n. 21, p. 2559–2560, 2008.

VAFG, M. **Análise de Rede Social no Facebook usando Gephi**. Disponível em: <<https://www.youtube.com/watch?v=zxQgaRUbl6g>>. Acesso em: 23 de mai. 2018.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração**. São Paulo: Atlas, 1997.

WATTS, D. J.; STROGATZ, S. H. **Collective dynamics of ‘small-world’ networks**. v. 393, n. June, p. 440–442, 1998.

WEKA. **Weka 3: Data Mining Software in Java**. Disponível em: <<https://www.cs.waikato.ac.nz/ml/index.html>>. Acesso em: 30 jun. 2018.