

MINERAÇÃO DE BUSCA DE SENTIMENTOS NAS REDES SOCIAIS

Autor: Jeferson Miranda

Prof. Orientador: Fernando César de Ornelas

Centro Universitário Uniasselvi de Brusque

Curso (SISTEMA DE INFORMAÇÃO - BACHARELADO) – Trabalho de Graduação

16/11/2022

RESUMO

O aumento das redes sociais nos últimos anos permitiu aos usuários se conectarem e compartilharem informações em tempo real, enviando-as a milhares de outros usuários em um curto espaço de tempo. Além disso, a maneira como os usuários interagem mudou. Os usuários de redes sociais costumam postar suas opiniões sobre os grandes eventos, lançamentos de produtos, catástrofes, epidemias, entre outros acontecimentos. Para acompanhar o que eles estão falando nas redes sociais pode ser um diferencial para as organizações que desejam elaborar melhores estratégias de marketing, obter feedback sobre algum produto ou determinado evento. Com a grande quantidade de dados crescendo dia a dia, a análise desses dados de forma não automatizada pode ser um problema enorme para a empresa. Como será abordado neste trabalho, o foco será apresentar os meios de comunicação e principais técnicas usada para coletar esses dados e as principais redes sociais.

Palavras-chave: Mineração de dados; Sentimentos; Redes sociais;

ABSTRACT

The rise of social networking in recent years has allowed users to connect and share information in real time, sending it to thousands of other users in a short amount of time. Also, the way users interact has changed. Users of social networks often post their opinions about major events, product launches, catastrophes, epidemics, among other events. To follow what they are talking about on social networks can be a differential for organizations that want to develop better marketing strategies, get feedback on a product or a certain event. With the large amount of data growing day by day, analyzing this data in a non-automated way can be a huge problem for the company. As will be discussed in this work, the focus will be to present the means of communication and the main techniques used to collect this data and the main social networks.

Keywords: Data mining; Feelings; Social networks;

1 INTRODUÇÃO

O rápido crescimento do uso da Internet e a popularização das redes sociais mudaram a forma de interação entre pessoas e organizações. Usuários publicam suas opiniões sobre as organizações, eventos, catástrofes, dentre outros, em seus perfis em redes sociais, que de maneira rápida são propagadas para vários outros usuários. Estas mensagens podem conter teores positivos, mas também podem ser duras críticas. Isso pode acarretar em vantagens para as organizações, mas também pode comprometer seriamente a imagem das mesmas, pelo enorme alcance de usuários que atualmente as redes sociais possuem. Essa popularização da Internet, por sua vez, gera um grande volume de informação a cada instante, e as organizações, em geral, não conseguem acompanhar no mesmo ritmo o que os usuários estão comentando sobre as mesmas. No entanto, percebeu-se que ao analisar essas informações, as organizações poderiam ter a vantagem de conhecer as opiniões dos usuários sobre seus serviços ou produtos fornecidos a partir de dados das redes sociais (Gomes, 2013).

Neste contexto, a Mineração de Textos, também conhecida como Descoberta de Conhecimento em Textos fornece um conjunto de técnicas que podem automatizar o processo de coleta e estruturação de informações e, junto com a Análise de Sentimentos, permite que as organizações possam saber o que os usuários estão comentando sobre elas em seus perfis na web. A partir dessa análise, as organizações podem se beneficiar dos resultados para os mais diversos fins, como elaborar estratégias de marketing, táticas de segurança, melhoria de serviços, dentre outros.

Reconhecer o estado emocional dos indivíduos por meio de suas postagens nas redes sociais pode ajudar a fortalecer estratégias para melhorar as partes defeituosas da organização, seja na gestão, na estrutura ou no ensino de instituições de ensino.

2 FUNDAMENTAÇÃO TEÓRICA

2.1.1 EVOLUÇÃO TECNOLÓGICA NAS REDES

Rede social como o próprio nome já sugere: “são todos os meios que proporcione a comunicação entre as pessoas”, mas essa definição genérica acaba englobando métodos nada tecnológicos, como o envio de cartas, por exemplo. Já nos

tempos modernos, rede social é como são chamados serviços que, usando a internet, permitem que as pessoas se conectem umas com as outras instantaneamente.

Na era pré-internet, o telégrafo fez as vezes de uma rede social pré-histórica, permitindo o envio de mensagens a longas distâncias usando a tecnologia. Então, antes da existência da internet como a conhecemos, existiu uma tecnologia militar chamada ARPANET, que conectava universidades em uma proto-internet. A primeira mensagem enviada por esta plataforma saiu da UCLA para o Stanford Research Institute em 1969 .(GNIPPER , 2018).

Figura 1 - Primeiro log IMP da ARPANET

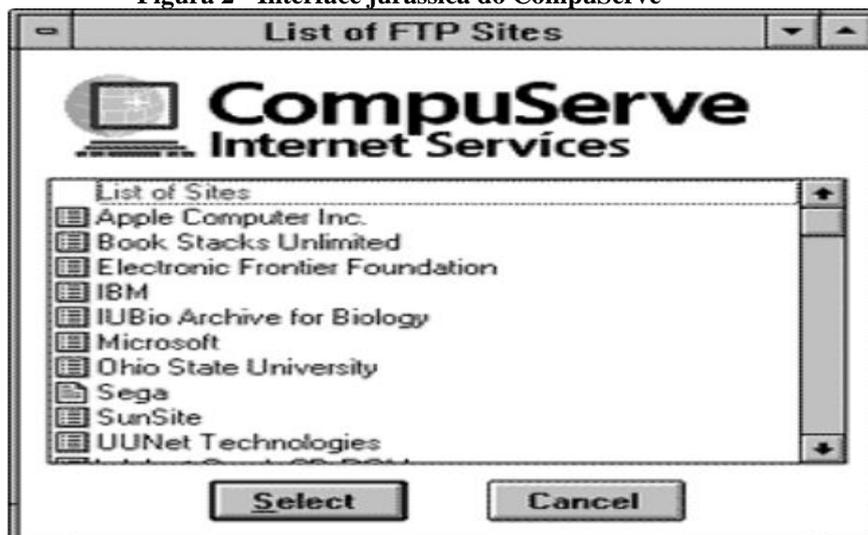
29 OCT 69	2100	LOADED OP. PROGRAM	CSK
		FOR BEN BARKER	
		BBV	
	22:30	Talked to SRI	CSK
		host to host	
		Left op imp program	CSK
		running after sending	
		a host dead message	
		to imp.	

FONTE: (CANALTECH,2018)

Na figura 1 é possível ver o primeiro log IMP da ARPANET mostra transcrição da primeira mensagem enviada pela plataforma, em 29/10/1969.

Então, já em 1982, a ARPANET foi a primeira rede a usar o protocolo TCP/IP, abrindo as portas para um mundo totalmente novo. Já com a chegada da computação pessoal, a CompuServe foi um dos primeiros serviços a oferecer conexão com a internet internacionalmente, sendo, também, a empresa que criou o formato GIF. Para acessar a rede, era preciso usar um cliente de acesso próprio, que oferecia serviços de e-mail e fóruns de discussão. Nos anos 1990, o serviço já era muito popular, sendo posteriormente fundido com a AOL, que dominou a internet na época, a figura 2 é possível ver a interface jurássica do CompuServe.

Figura 2 - Interface jurássica do CompuServe



FONTE: (CANALTECH,2018)

Ainda que timidamente, essas novidades começaram a mudar a forma com que as pessoas se comunicam, permitindo, ainda, conhecer gente que não conheceríamos ao vivo por lá. Isso já marcou o início da transformação da nossa sociedade, com as pessoas cada vez mais buscando a internet como meio de entrar em contato com outras, se informar e se entreter (GNIPPER, 2018).

Nos últimos anos, o *boom* das mídias sociais influenciou diretamente inúmeras empresas a se renderem ao mundo digital. O fato é que, diariamente, novos perfis corporativos são criados em diversas redes sociais, porém poucas são as empresas que realmente entendem a importância e a dinâmica das redes sociais (OLIVEIRA, 2013).

A velocidade de propagação das informações geradas se dá nos dois casos, tanto nos negativos, quanto nos positivos. Com isso, outra oportunidade que surgiu foi a possibilidade de monitorar as conversas dos internautas, a fim de antecipar tendências e avaliar como está seu nicho de mercado. Medir o que o consumidor fala sobre a concorrência, fundamentando a sobrevivência a um mercado com clientes cada vez menos fiéis e cada vez mais ávidos por empresas transparentes e socialmente responsáveis (GRIGOLETTI, 2011).

Monitorar as redes sociais, pensando em estratégias e executando ações para divulgar produtos e serviços, pesquisar público-alvo e novas vertentes para o negócio e, claro, responder a questões específicas de internautas, especialmente quando são críticas à empresa. Essas são as funções básicas de um analista de mídias sociais, profissional cada vez mais requisitado por companhias de todas as

áreas de atuação, que apostam na velocidade da web e na expansão das redes sociais para alavancar seus ganhos (KOPPSCHITZ, 2011).

Conforme Liu (2010), a informação textual pode ser classificada em dois tipos principais: fatos e opiniões. Os fatos são expressões objetivas sobre entidades, eventos e suas propriedades, já opiniões são geralmente expressões subjetivas que descrevem os sentimentos e avaliações das pessoas sobre determinada entidade.

Palavras de opinião são palavras comumente usadas para expressar sentimentos negativos e positivos. Por exemplo, “belo, bom, maravilhoso e incrível” são palavras de opiniões positivas, e “ruim, pobre e terrível” são palavras de opiniões negativas. Frases e palavras de opiniões são fundamentais para análise de sentimentos por razões óbvias (LIU, 2010). Com isso, pode-se conceituar que o processamento de linguagem natural é a compreensão de uma informação falada ou escrita por meio de regras equivalentes àqueles existentes na comunicação linguística humana. A fim de melhor compreender como ocorre o processamento de uma informação textual, deve-se analisar as fases do processamento de linguagem natural, que são: Análise Morfológica, Análise Sintática, Análise Léxica, Análise Semântica e Análise Pragmática (CESÁRIO JÚNIOR, 2010).

A área de mineração de texto é fortemente baseada na área de mineração de dados e KDD (*Knowledge Discovery in Databases*), sendo, por isso, conhecida também como *Text Data Mining* (HEARST, M. A., 1999 *apud* SOUZA; LINDGREN, 2011) e KDT (*Knowledge Discovery in Texts*). A mineração de texto, assim como a Mineração de Dados, é também interdisciplinar com influências de áreas como Processamento de Linguagem Natural, Recuperação de Informação, Inteligência Artificial e Ciência Cognitiva. A conjugação do conhecimento dessas áreas fez da mineração de texto uma área própria, chamada apenas de Mineração de Texto.

2.1.2 REDES SOCIAIS

As redes sociais: são espaços de relacionamento, lucrativas, permitem a coleta de dados e relatórios relevantes, métricas. Possibilitam a presença em mais um canal de comunicação, além de ampliar o alcance, a popularidade e, dependendo da estratégia, pode melhorar ou piorar a reputação de uma empresa.

Figura 3 - Ícones redes



FONTE: <https://superatualizado.com.br/redes-sociais/>

É importante lembrar que reputação é algo que uma vez perdida, dificilmente será recuperada. Portanto, é fundamental saber da importância das redes sociais para empresas, mas também de uma gestão de mídias assertiva (GIOMARÃES FILHO,2018).

Podendo também fazer uma análise mais profunda do perfil de seus clientes e definir com muito mais precisão qual é o seu público ideal. Elas integram diretamente as estratégias de marketing, mas impactam à empresa como um todo (GRANDCHAMP,2022).

Dentre as diversas mídias sociais existentes, será realizado uma análise sobre algumas ferramentas disponibilizadas para mineração de dados, usando como exemplo em com a duas mídias social uma delas é o Twitter e a outra o Facebook e, hoje chamada também de META, que também e dona de outras três redes sociais (WhatsApp, Instagram e Messenger), que são muito utilizadas, onde as mesmas aparece entre as 6 do rank das 10 mais utilizadas no brasil e mundo (VOLPATO,2022).

2.1.3 FACEBOOK

O Facebook é uma rede social que teve seu início em 2003. Os estudantes de Harvard Mark Zuckerberg, Dustin Moskovitz e Chris Hughes criaram uma plataforma de comunicação destinada a outros alunos da universidade. Logo o

serviço se expandiu para outras instituições de ensino. Foi em 2004 que Zuckerberg e seus parceiros lançaram o site thefacebook.com, já totalmente aberto ao público. Em 2005, a empresa abandona de vez o “the” em seu nome e passa a ter o nome que a consagrou. Na figura 4 é possível ver a primeira versão do “TheFacebook”.

Figura 4 – PRIMEIRA VERSÃO DO “THEFACEBOOK”



FONTE: TECNOBLOG.NET

Não demorou muito para que o Facebook desbancasse outras redes sociais e se tornasse a mais popular do mundo. Em 2008, a rede social já era a mais popular no mundo, mas a virada definitiva aconteceu em 2009, quando tomou o lugar do MySpace como a mais usada nos Estados Unidos, com mais de 130 milhões de usuários.

Com o tempo, o Facebook se estabeleceu e é a maior rede social do mundo com folga, com quase 3 bilhões de usuários ativos mensais em 2022. Na plataforma, usuários podem criar seus perfis pessoais ou páginas que representam organizações ou celebridades. O “like”, ou “curtida”, com um polegar para cima é um dos símbolos maiores da rede social.

Um dos recursos mais marcantes da rede social é o feed de notícias, uma área com curadoria de conteúdo por meio de algoritmo que mostra posts de amigos, contatos e páginas curtidas. O site também pode exibir publicações que o sistema

julgar interessante ao usuário, mas de perfis e empresas que ele não conhecia previamente.

Figura 5 - FUNDADORES FACEBOOK



FONTE: TECNOBLOG.NET

Para acesso aos recursos de análise de texto na plataforma, o FACEBOOK possui uma API chamada a Graph API é a principal forma de os aplicativos lerem e gravarem no gráfico social do Facebook (FACEBOOK, 2022).

Abaixo uma visão geral disponível na documentação Graph API.

2.1.4 TWITTER

Com mais de 600 milhões de usuários cadastrados, o Twitter é uma rede social que permite a seus usuários postarem mensagens de texto rápidas, limitadas a 140 caracteres, conhecidas como tweets. Sua estrutura dinâmica permite que qualquer usuário tenha acesso às informações que são constantemente postadas, sem que para isso restrinja aos usuários possuírem alguma permissão de conexão entre eles. Russel (2013) aponta que esse é o grande diferencial do Twitter em relação a outras redes

sociais populares, como o LinkedIn e o Facebook, tornando o Twitter uma rede social mais interessante de se explorar.

O Twitter fornece uma REST API para desenvolvedores, que permite aos mesmos acessarem atualizações, status, dados de usuários, etc. REST (Representation State Transfer) é uma arquitetura de redes de estilo híbrido derivada do estilo de várias arquiteturas baseadas em rede, que define uma interface conectora que permite aos clientes “conversar” com servidores de maneira única. O Twitter também fornece aos desenvolvedores acesso a um grande volume de informações em tempo real através de uma Streaming API. A Twitter Streaming API fornece para desenvolvedores acessos de baixa latência para a stream global do Twitter de dados de tweets.

O Twitter fornece uma REST API para desenvolvedores, que permite aos mesmos acessarem atualizações, status, dados de usuários, etc. REST (Representation State Transfer) é uma arquitetura de redes de estilo híbrido derivada do estilo de várias arquiteturas baseadas em rede, que define uma interface conectora que permite aos clientes “conversar” com servidores de maneira única. O Twitter também fornece aos desenvolvedores acesso a um grande volume de informações em tempo real através de uma Streaming API. A Twitter Streaming API fornece para desenvolvedores acessos de baixa latência para a stream global do Twitter de dados de tweets.

Os tweets podem ser agrupados por hashtags palavras precedidas pelo caractere #, utilizado para marcar palavras-chave ou tópicos em um tweet. 5 Além disso, os usuários podem dar retweet em um tweet publicado por outros usuários. Um retweet é uma nova postagem do tweet de alguém. O autor deste trabalho manteve os retweets para análise, por entender que quando um usuário “retuita” um tweet de alguém, ele está concordando com a opinião de quem postou o tweet.

O Twitter foi utilizado como fonte de exploração de dados para este trabalho, tendo em vista que o Twitter é uma rede social com uma grande quantidade de usuários ativos e possui alcance global, que induz seus usuários a compartilhar suas ideias constantemente, gerando grande quantidade de informação a cada instante, além de fornecer APIs que permitem explorar dados dos tweets de maneira fácil.

Rodrigues Barbosa et al (2012) pontuam que o modelo de interação do Twitter induz os usuários a compartilhar e expressar continuamente suas opiniões e sentimentos, que são propagados para seus respectivos seguidores. Porém,

determinar o sentimento que cada tweet expressa pode ser uma tarefa trabalhosa, sujeita a erros e ambiguidades. Para contornar esses desafios, Rodrigues Barbosa et al (2012) exploram hashtags. As hashtags são utilizadas para determinar o sentimento expresso pelos usuários do Twitter nos tweets referentes às eleições presidenciais do Brasil, em 2010.

A classificação das hashtags em algum sentimento que indique sua polaridade é realizada de maneira manual. Russel (2013) cita a curiosidade humana e a necessidade de compartilhar ideias e experiências, fazer perguntas, interagir de maneira rápida. O Twitter propicia de maneira dinâmica que todos esses aspectos sejam possíveis de serem realizados na velocidade do pensamento. Além disso, a rede social possui um diferencial das demais redes sociais, pelo seu modelo assimétrico de seguidores, onde qualquer usuário pode ficar por dentro dos últimos acontecimentos, mesmo que ele não siga o autor da postagem, enquanto em outras redes sociais, como o Facebook e LinkedIn, é preciso uma aceitação de conexão entre os seus usuários (RUSSEL, 2013).

Tendo em vista a grande vantagem de utilizar o Twitter como rede social para mineração, este trabalho busca detectar as opiniões dos usuários do Twitter.

2.1.5 INSTAGRAM

O Instagram foi criado por Kevin Systrom e pelo brasileiro Mike Krieger em 2010. Em poucos meses, a rede social se tornou um dos aplicativos mais promissores da App Store.

Em apenas um ano, o Instagram já contava com dez milhões de usuários, mesmo limitado ao ecossistema da Apple e disponível apenas nos iPhones. Apenas em 2012 foi lançada uma versão para o Android, mas só depois da aquisição pelo Facebook por US\$ 1 bilhão.

Inicialmente, o aplicativo ganhou popularidade com um formato simples, que previa a publicação de imagens em proporção quadrada com uma seleção de filtros artísticos. Com o tempo, o app derrubou algumas de suas restrições e liberou o compartilhamento de fotos e vídeos em formatos mais variados.

Em 2016, pressionado pelo crescimento do Snapchat, o Instagram lança os Stories, formato de publicação efêmera, com vídeos de até 15 segundos e que

desaparecem após 24 horas. O lançamento foi bem recebido e abriu portas para mais ferramentas, como transmissões de vídeo ao vivo.

Em 2021, o Instagram superou a marca de 2 bilhões de usuários ativos por mês, marca que o coloca na posição de segunda maior rede social do mundo.

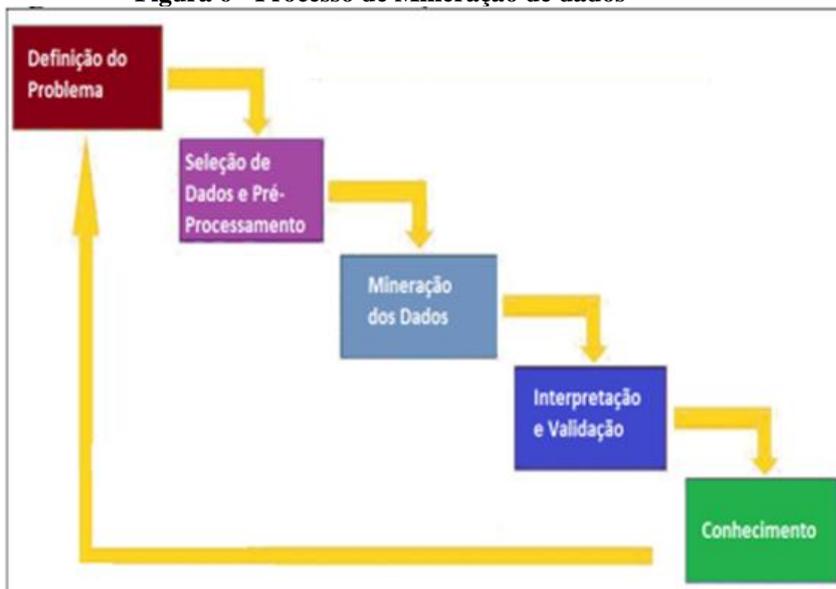
(CANALTEC, 2022)

2.2 MINERAÇÃO DE DADOS

A Mineração de Dados, também conhecida como Descoberta de Conhecimento em Banco de Dados, foca na exploração computadorizada em grandes massas de dados para descobrir padrões interessantes entre elas (FELDMAN et al, 1998). A maioria dos trabalhos de mineração de dados é realizada em cima de base de dados estruturada.

Segundo Moraes e Ambrósio (2007), é importante que os resultados do processo de Descoberta de Conhecimento em Banco de Dados sejam compreensíveis para nós humanos, e principalmente para os usuários finais do processo, que são geralmente pessoas que tomam decisões nas organizações. Os processos de Mineração de Dados (Figura 6) devem ser vistos como práticas para melhorar os resultados das explorações feitas utilizando ferramentas tradicionais de exploração de dados. O processo de Descoberta de Conhecimento em Banco de Dados se assemelha ao de Descoberta de Conhecimento em Textos, desde a definição do problema, até a extração de conhecimentos. Na definição do problema, os objetivos a serem alcançados através da DCBD são identificados. Na seleção de dados e pré-processamento, os dados normalmente não estão em um formato adequado para extração de conhecimento, por isso faz-se necessária a aplicação de métodos para extração e integração, transformação, limpeza, seleção e redução de volume destes dados, antes da etapa de mineração (MORAIS; AMBRÓSIO, 2007). Na fase de mineração de dados, busca-se cumprir os objetivos determinados na fase de definição de problema. É nessa etapa que são configurados e executados algoritmos em busca de padrões relevantes, podendo executá-los diversas vezes, até que sejam alcançados os resultados esperados. Na etapa de conhecimento, diversos padrões podem ser identificados, alguns não relevantes, outros interessantes para o domínio do problema.

Figura 6 - Processo de Mineração de dados



Fonte: Elaborado pelo autor.

2.3 MINERAÇÃO DE TEXTOS

“A Mineração de Textos, também conhecida como Descoberta de Conhecimento de Texto refere-se ao processo de extrair padrões interessantes e não triviais ou conhecimento a partir de textos desestruturados.” (TAN, 1999, p.1, tradução livre).

A Mineração de Textos aplica as mesmas funções analíticas da Mineração de Dados (GOMES, 2013), porém para dados textuais. Segundo Hearst (1999), os dados textuais englobam uma vasta e rica fonte de informação, mesmo em um formato que seja difícil de extrair de maneira automatizada. Esta grande massa de informação textual não estruturada não pode ser utilizada por computadores para extração de conhecimento, uma vez que os mesmos a tratam apenas como uma sequência de caracteres. Assim, faz-se necessária a aplicação de diferentes métodos e algoritmos para dar estruturação aos dados textuais, visando facilitar a extração de conhecimento dos respectivos dados. Embora a Mineração de Textos geralmente se refira à extração de conhecimento de base de dados textuais, pode-se recorrer a outras áreas para facilitar o processo de descoberta de conhecimento nos dados textuais.

Figura 7 - Etapas do processo de Mineração de Textos



Fonte: Aranha (2007)

Aranha (2007) propõe um modelo de Mineração de Textos (Figura 7), contendo cinco fases distintas: coleta, pré-processamento, indexação, mineração e análise. Na coleta, utiliza-se web crawlers, programas que visitam sítios e extraem informações, para extração dos textos que serão utilizados para a extração de conhecimento. No pré-processamento, são utilizadas técnicas como o Processamento de Linguagem Natural para estruturar os textos que serão analisados. A indexação é a etapa onde são extraídos conceitos dos documentos através da análise de seu conteúdo e traduzidos em termos da linguagem de indexação. Esta representação identifica o documento e define seus pontos de acesso para consultas (GOMES, 2006). Na etapa de mineração, são aplicados métodos e algoritmos para a identificação de padrões interessantes e extração de conhecimento. Na parte de análise, os resultados são avaliados e validados.

Os passos de execução deste trabalho são baseados no processo de Mineração de Textos proposto por Aranha (2007), com exceção do passo de indexação, pois não foi considerado relevante para a execução deste trabalho, pois não houve necessidade de realizar consultas nos textos que compõem a base de dados.

2.4 MINERAÇÃO DE TEXTOS PROCESSAMENTO DA LINGUAGEM NATURAL

Chama-se Processamento de Linguagem Natural (PLN) um conjunto de técnicas teórico-computacionais que visam representar dados textuais e processar a linguagem natural humana para diversas tarefas.

O Processamento da Linguagem Natural é um conjunto de técnicas computacionais para analisar e representar ocorrências naturais de texto em um ou mais níveis de análise linguística com o objetivo de se alcançar um processamento de linguagem similar ao humano para uma série de tarefas ou aplicações. (LIDDY, 2001, p.1, tradução livre).

O PLN naturalmente lida com diversos elementos linguísticos e estrutura gramatical, sendo um processo complexo, paralelo à complexidade da linguagem natural. Liddy (2001) mostra que para processar a linguagem natural, o PLN a representa em diversos níveis, como léxico, morfológico, semântico, etc.

Para auxiliar na etapa de pré-processamento de dados textuais, técnicas de PLN podem ser utilizadas, como remoção de stopwords, segmentação de palavras, lematização, dentre outras. Stopwords são palavras muito comuns que aparecem no texto e carregam pouco significado; servem apenas com uma função sintática, mas não indicam importância ao assunto (EL-KHAIR, 2006). Alguns exemplos de stopwords: as, e, os, de, para, com.

Tendo em vista a estruturação dos dados a serem utilizados nesse trabalho para melhorar o procedimento de classificação de texto, a PLN fornece técnicas que podem ser utilizadas para melhor estruturar os dados que serão utilizados neste trabalho, como por exemplo, removendo palavras dos tweets que serão minerados que não expressam algum sentimento ou polaridade, como artigos, preposições, dentre outros elementos gramaticais.

2.5 NAIVE BAYES

O Naive Bayes é um simples classificador probabilístico baseado na aplicação do teorema de Bayes. Ele é frequentemente utilizado como base na classificação de textos por ser rápido e fácil de implementar (RENNIE et al, 2003, p. 1, tradução livre). Gomes (2013) cita que o classificador Naive Bayes é considerado um dos mais eficientes em questões relacionadas com processamento e precisão na classificação de novas amostras.

Figura 8 - Teorema de Bayes

$$P(A|B) = \frac{P(B_1|A).P(B_2|A)...P(B_n|A).P(A)}{P(B_1).P(B_2)...P(B_n)}$$

Fonte Elaborado pelo autor

A Figura 8 ilustra o teorema de Bayes. Assumindo que B representa um evento que ocorreu previamente e A um evento que depende de B, para que seja calculada a probabilidade de A ocorrer dado o evento B, o algoritmo deverá contar o número de casos em que A e B ocorrem juntos e dividir pelo número de casos em que B ocorre sozinho.

O algoritmo bayesiano utilizado na etapa de Mineração de Textos neste trabalho é uma implementação pertencente ao Apache Mahout. O Mahout é uma biblioteca de aprendizagem de máquina de código aberto do Apache (OWEN,2011).

2.6 ANÁLISE DE SENTIMENTOS

A larga expansão da internet gera muitas informações em forma de opiniões em seus diversos canais: fóruns, comunidades, redes sociais, etc. INDURKHYA e DAMERAU (2010) citam que as opiniões são tão importantes que, onde quer que se queira tomar decisões, as pessoas querem ouvir a opinião de outros. Isso não é uma verdade apenas para as pessoas, como também para as organizações, afinal, conhecer a opinião dos clientes acerca dos seus produtos e serviços é de grande valia para as organizações.

A Análise de Sentimentos ou Mineração de Opinião é o estudo computacional de opiniões, sentimentos e emoções expressos em texto. A informação textual pode ser classificada em dois principais tipos: fatos e opiniões. Fatos são expressões objetivas sobre entidades, eventos e suas propriedades. Opiniões são geralmente expressões subjetivas que descrevem os sentimentos da população, avaliações, ou sentimentos em relação às entidades, e suas propriedades (INDURKHYA; DAMERAU, 2010). Ela é a área que ajuda de maneira automatizada a determinar a direção do sentimento em textos (positivo ou negativo).

Gomes (2013) pontua que apesar da Análise de Sentimentos ser apresentada por grande parte da literatura como estudo computacional de sentimentos, a mesma pode ser utilizada para muitos outros projetos. Tendo em vista que a Análise de Sentimentos se trata de um problema de classificação, ela pode ser utilizada para classificar dados textuais, segundo sua polaridade, mesmo se o texto não denotar algum sentimento.

3 MATERIAL E MÉTODOS

Neste capítulo será descrito os métodos utilizados para chegar ao objetivo do trabalho foi a Bibliográfica utilizado de material já publicado, constituído basicamente de livros, artigos de periódicos e, atualmente, de informações disponibilizadas na internet.

Bem como algumas das ferramentas que pode ser utilizada para o desenvolvimento do mesmo.

3.1 FERRAMENTAS

As ferramentas pesquisadas para o desenvolvimento, mineração de textos e análise de sentimentos foram as seguintes:

- Linguagem de programação Python
- Colaboratory - Escreva e execute código Python arbitrário pelo navegador
- API do FaceBook

- Biblioteca para utilizar com Python para análise.
 - Requests
 - Pandas
 - Numpy
 - Plotly.express

3.2 LINGUAGEM DE PROGRAMAÇÃO PYTHON

A linguagem Python segundo (SUMMERFIELD, 2012) é bastante popular, por ser simples para ler e escrever; é uma linguagem multiplataforma e que pode ser usado para programar em paradigma procedural ou orientada a objetos. Possui uma completa biblioteca- padrão que permite inúmeras possibilidades com uma ou poucas linhas de código. Lançada por Guido van Rossum em 1991 como software livre, pode ser utilizada gratuitamente.

Segundo o site <http://www.tiobe.com/tiobe-index/>, a linguagem Python cresceu no ranking das linguagens mais utilizadas em 2021 ficando em primeiro lugar.

A linguagem Python foi utilizada para demonstra sua utilização na extração dos dados, remover caracteres especiais e espaços em branco do arquivo extraído entre outras funcionalidade que a linguem permite.

3.3 COLABORATORY

O Colaboratory ou “Colab” é um produto do Google Research, área de pesquisas científicas do Google. O Colab permite que qualquer pessoa escreva e execute código Python arbitrário pelo navegador e é especialmente adequado para aprendizado de máquina, análise de dados e educação. Mais tecnicamente, o Colab é um serviço de notebooks hospedados do Jupyter que não requer nenhuma configuração para usar e oferece acesso sem custo financeiro a recursos de computação como GPUs. (GOOGLE RESEARCH, 2022)

3.4 API do FACEBOOK

A Graph API é a principal forma de inserir e retirar dados da plataforma do Facebook. Trata-se de uma API baseada em HTTP que os aplicativos podem usar para consultar dados programaticamente, publicar novas histórias, gerenciar anúncios, carregar fotos e realizar uma ampla variedade de tarefas.

A Graph API tem esse nome com base na ideia de um "gráfico social" — uma representação das informações no Facebook. O gráfico social é composto por nós, bordas e campos. Em geral, os nós são usados para a obtenção de dados sobre um objeto específico, as bordas são usadas para obter coleções de objetos sobre um objeto único e os campos para obter dados sobre um objeto único ou sobre cada objeto de uma coleção (FACEBOOK, 2022).

3.4.1 HTTP

Todas as transferências de dados estão em conformidade com o HTTP/1.1, e todos os pontos de extremidade exigem HTTPS. Como a Graph API se baseia em

HTTP, ela funciona com qualquer linguagem que tenha uma biblioteca desse tipo, como cURL e urllib. Isso quer dizer que você pode usar a Graph API diretamente em seu navegador. Por exemplo, solicitar esta URL em seu navegador...

```
curl -i -X GET  
  
"https://graph.facebook.com/facebook/picture?redirect=false"
```

3.4.2 URL DE HOSPEDAGEM

Quase todas as solicitações são enviadas à URL de hospedagem graph.facebook.com.

A única exceção são os carregamentos de vídeo, que usam o graph-video.facebook.com.

3.4.3 TOKENS DE ACESSO

Os tokens de acesso permitem que seu aplicativo acesse a Graph API. Todos os pontos de extremidade da Graph API exigem algum tipo de token de acesso. Por isso, sempre que você acessar um ponto de extremidade, sua solicitação deverá conter essa informação. Normalmente, eles desempenham duas funções:

- Permitem que seu aplicativo acesse as informações de um usuário sem exigir a senha dele.
- Permitem que identifiquemos seu aplicativo, o usuário que está usando ele e quais dados o usuário autorizou o aplicativo a acessar.

3.4.4 NÓ

Um nó é um objeto individual com um ID único. Por exemplo, existem vários objetos de nós de usuário, cada um com um ID único representando uma pessoa no Facebook. Páginas, Grupos, Publicações, Fotos e Comentários são apenas alguns dos nós do gráfico social do Facebook.

O exemplo cURL a seguir representa uma chamada ao nó do usuário.

```
curl -i -X GET \

  "https://graph.facebook.com/USER-ID?access_token=ACCESS-
  TOKEN"
```

Por padrão, essa solicitação retorna os seguintes dados formatados usando JSON:

```
{

  "name": "Your Name", "id": "YOUR-USER-ID"

}
```

3.4.5 METADADOS DO NÓ

É possível obter uma lista de todos os campos de um objeto de nó (como usuário, página ou foto), incluindo o nome, a descrição e o tipo de dado do campo. Envie uma solicitação GET para um ID de objeto e inclua o parâmetro metadata=1:

```
curl -i -X GET \

  "https://graph.facebook.com/USER-ID?

  metadata=1&access_token=ACCESS-TOKEN"
```

A resposta JSON incluirá a propriedade metadata, que lista todos os campos compatíveis para o nó específico:

```
{ "name": "Jane Smith",

  "metadata": { "fields": [ {

    "name": "id",
```

```

    "description": "The app user's App-Scoped User ID. This ID is unique to the app and cannot be used by other apps.",
    "type": "numeric string"
  },
  {
    "name": "age_range",
    "description": "The age segment for this person expressed as a minimum and maximum age. For example, more than 18, less than 21.",
    "type": "agerange"
  },
  {
    "name": "birthday",
    "description": "The person's birthday. This is a fixed format string, like `MM/DD/YYYY`. However, people can control who can see the year they were born separately from the month and day so this string can be only the year (YYYY) or the month + day (MM/DD)",
    "type": "string"
  },
},

```

3.4.6 /ME

O nó /me é um ponto de extremidade especial que se traduz no ID do objeto da pessoa ou Página cujo token de acesso está sendo usado para fazer as chamadas à API. Se você tiver um token de acesso do usuário, poderá recuperar o nome e a ID desse usuário ao usar o seguinte:

```
curl -i -X GET \
```

```
"https://graph.facebook.com/me?access_token=ACCESS-TOKEN"
```

3.4.7 BORDAS

Uma borda é uma conexão entre dois nós. Por exemplo, um nó de usuário pode estar conectado a fotos, e um nó de foto pode estar conectado a comentários. O exemplo cURL a seguir retornará uma lista das fotos publicadas por uma pessoa no Facebook.

```
curl -i -X GET \  
  "https://graph.facebook.com/USER-  
  ID/photos?access_token=ACCESS-TOKEN"
```

Cada ID retornada representa um nó de foto e quando ela foi carregada no Facebook.

```
{  
  "data": [  
    {  
      "created_time": "2017-06-06T18:04:10+0000",  
      "id": "1353272134728652"  
    },  
    {  
      "created_time": "2017-06-06T18:01:13+0000",  
      "id": "1353269908062208"  
    }  
  ],  
}
```

3.4.8 CAMPOS

Os campos são propriedades dos nós. Ao consultar um nó ou uma borda, um conjunto de campos será retornado por padrão, conforme os exemplos acima. No entanto, é possível especificar que campos você deseja retornar usando o parâmetro `fields` e listando cada campo. Isso substitui os padrões e retorna somente os campos especificados, além do ID do objeto, que sempre é retornado.

A solicitação cURL a seguir inclui o parâmetro `fields` e o nome, o email e a foto de perfil do usuário.

```
curl -i -X GET \  
  
  "https://graph.facebook.com/USER-  
ID?fields=id,name,email,picture&access_token=ACCESS-TOKEN"
```

Dados retornados

```
{  
  "id": "USER-ID",  
  "name": "EXAMPLE NAME",  
  "email": "EXAMPLE@EMAIL.COM",  
  "picture": {  
    "data": {  
      "height": 50,  
      "is_silhouette": false,  
      "url": "URL-FOR-USER-PROFILE-PICTURE",  
      "width": 50  
    }  
  }  
}
```

3.5 BIBLIOTECA REQUESTS

Requests permite enviar solicitações HTTP/1.1 com extrema facilidade. Não há necessidade de adicionar strings de consulta manualmente às suas URLs ou codificar seus PUT& POST dados por formulário — mas hoje em dia, basta usar o json método. (PAYTON, 2022)

3.6 BIBLIOTECA PANDAS

Pandas é uma ferramenta de análise e manipulação de dados de código aberto rápida, poderosa, flexível e fácil de usar, construída sobre a linguagem de programação Python. (PANDAS, 2022)

3.7 BIBLIOTECA NUMPY

NumPy é o pacote fundamental para computação científica em Python. É uma biblioteca Python que fornece um objeto de matriz multidimensional, vários objetos derivados (como matrizes e matrizes mascaradas) e uma variedade de rotinas para operações rápidas em matrizes, incluindo matemática, lógica, manipulação de forma, classificação, seleção, E/S , transformadas discretas de Fourier, álgebra linear básica, operações estatísticas básicas, simulação aleatória e muito mais. (NUMPY, 2022)

3.8 BIBLIOTECA PLOTLY.EXPRESS

O plotly express módulo (geralmente importado como px) contém funções que podem criar figuras inteiras de uma só vez e é conhecido como Plotly Express ou PX. Plotly Express é uma parte integrada da plotly biblioteca e é o ponto de partida recomendado para criar as figuras mais comuns. Toda função Plotly Express usa objetos gráficos internamente e retorna uma plotly.graph_objects. Figure instância. Ao longo da plotly documentação, você encontrará o modo Plotly Express de

construir figuras no topo de qualquer página aplicável, seguido por uma seção sobre como usar objetos gráficos para construir figuras semelhantes. Qualquer figura criada em uma única chamada de função com o Plotly Express pode ser criada usando apenas objetos gráficos, mas com entre 5 e 100 vezes mais código. (PLOTLY, 2022)

4 CONCLUSÃO

Este trabalho tem como objetivo pesquisar ferramenta que auxilie analistas de mídias sociais a identificar os sentimentos dos comentários inseridos nessas redes, e avaliar o grau de impacto e proporcionar aos analistas maior agilidade no atendimento ao consumidor.

Foi realizada leitura de vários trabalhos referentes a análise de sentimentos que abordam diversas formas de chegar ao resultado esperado. Todos os trabalhos que foram referência ao trabalho proposto usavam mineração de textos. A presente proposta, além de ater-se e atender aos conceitos de mineração de dados.

As estas três redes sociais Facebook, Instagram, Twitter é uma base de dados fortemente estudada na aplicação de mineração de dados, como análise de sentimentos e/ou opinião.

O fato de as empresas locais não possuírem um grande fluxo de informações no Facebook, Instagram, Tweets e as grandes empresas contatadas não permitirem o acesso direto ao aplicativo, o que traz complicações para o teste da ferramenta, que envolve a captura manual de dados para a realização do teste

Em termos de análise sentimentos, dois fatores se intensificaram no desenvolvimento: um é a alta taxa de erros de escrita do português, e o outro é que há muita ironia nos comentários, deixando grande parte dos comentários que precisam ser analisados corretamente.

A pesquisas levam em consideração que algumas redes sociais tem um vasto vocabulário da língua portuguesa, o grande número de palavras ambíguas e os erros de português.

Este trabalho mostrou algumas ferramentas e maneiras para extrair dados utilizando os modelos matemáticos e algoritmos, com a linguagem Python e bibliotecas específicas para conexão, nas APIs disponibilizadas para os

desenvolvimentos e análise das redes sociais citadas neste artigo para extração dos dados.

Para trabalhos futuros, é possível implementar as etapas e recursos apresentado neste trabalho, se aprofundar mais na linguagem python e suas bibliotecas para buscar melhor recurso da aplicação.

REFERÊNCIAS

- ARANHA, C.N. **Uma Abordagem de Pré-Processamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional**. 2007. 144 f. Tese (Doutorado em Engenharia Elétrica) — Pontífica Universidade Católica do Rio de Janeiro, Rio de Janeiro. 2007.
- CANALTECH. 06 Fervereiro 2018. Disponível em: <<https://canaltech.com.br/redes-sociais/a-evolucao-das-redes-sociais-e-seu-impacto-na-sociedade-parte-1-107830/>>. Acesso em: 20 Outubro. 2022.
- CANALTEC. **Instagram** Disponível em: <<https://canaltech.com.br/empresa/instagram/>>. Acesso em: 16 Novembro. 2022.
- CESÁRIO JÚNIOR . J.M. **Sobre o Conceito de Processamento de Linguagem Natural**, Tese de Mestrado (Tecnologia), Limeira SP, UNICAMP FT, 2010.
- EL-KHAIR, Ibrahim Abu. **Effects of stop words elimination for Arabic information retrieval: a comparative study**. *International Journal of Computing & Information Sciences*, v. 4, n. 3, p. 119-133, 2006. MAN, Ronen et al. Knowledge Management: A Text Mining Approach. In: PAKM.
- FACEBOOK . 2022. Disponível em: <<https://developers.facebook.com/docs/graph-api/overview>>. Acesso em: 20 Outubro. 2022.
- FELDMAN , Ronen et al. **Knowledge Management: A Text Mining Approach**. In: **PAKM**. 1998
- GIOMARÃES FILHO, P.H. **A importância das redes sociais para empresas**, 23 Janeiro. 2018. Disponível em: <<https://superatualizado.com.br/redes-sociais/>> Acesso em: 20 Outubro. 2022.
- GNIPPER . P. **A evolução das redes sociais e seu impacto na sociedade – Parte 1**, 06 Fevereiro 2018. Disponível em: <<https://canaltech.com.br/redes-sociais/a-evolucao-das-redes-sociais-e-seu-impacto-na-sociedade-parte-1-107830/>> Acesso em: 20 Outubro. 2022.
- GOMES, G. R. R. **Integração de Repositórios de Sistemas de Bibliotecas Digitais e de Sistemas de Aprendizagem**. 2006. 143 f. Tese (Doutorado em Informática) — Pontífica Universidade Católica do Rio de Janeiro, Rio de Janeiro. 2006.

GOMES, Helder Joaquim Carvalheira. **Text Mining: análise de sentimentos na classificação de notícias. Information Systems and Technologies (CISTI)**, 2013 8th Iberian Conference on. Lisboa. 2013.

GOGONI, R. **Quem fundou o Facebook?** ,2018. Disponível em: <<https://tecnoblog.net/responde/quem-fundou-o-facebook/>> Acesso em: 20 Outubro. 2022.

GOOGLE Research. 2022. **Noções básicas Colaboratory** Disponível em: <<https://research.google.com/colaboratory/intl/pt-BR/faq.html>>. Acesso em: 16 Novembro. 2022.

GRANDCHAMP, L **A importância das redes sociais para as empresas**, 19 Mai. 2022. Disponível em: <https://www.jornalcontabil.com.br/entenda-a-importancia-das-midias-sociais-para-empresas/>> Acesso em: 20 Outubro. 2022.

GRIGOLETTI, E. **Redes Sociais: O que mudou nas empresas?** 19 jul. 2011. Disponível em:<<http://www.clientesa.com.br/artigo/42951/redes-sociais-o-que-mudou-nas-empresas/ler.aspx>> Acessoem: 20 outubro 2022.

HEARST, M. A. Untangling text data mining. **Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics** (pp. 3— 10),1999. Stroudsburg, PA, USA: Association for Computational Linguistics.

KOPSCHITZ, I. **Analista de redes sociais, um profissional do presente - Jornal O Globo** , 27 set. 2011. Disponível em <<https://oglobo.globo.com/economia/emprego/analista-de-redes-sociais-um-profissional-do-presente-2747776>>. Acesso em: 20 Outubro. 2022.

LIU, B. Sentiment Analysis and Subejctivity. In Nitin Indurkha & F. J. Damerau (Eds.). **Handbook of Natural Language Processing Second Edition**. Boca Raton, FL: CRC Press, Taylor and Francis Group, Chicago: 2010 cap. 1, p.1 - 5.

LIDDY, E. **Natural Language Processing. Encyclopedia of Library and Information Science**. New York: Marcel Decker, Inc, 2001 LIU, B. **Sentiment Analysis and Subejctivity. In Nitin Indurkha & F. J. Damerau (Eds.). Handbook of Natural Language Processing Second Edition**. Boca Raton, FL: CRC Press, Taylor and Francis Group, Chicago: 2010 cap. 1, p.1 - 5.

MORAIS, Edilson Andrade Martins; AMBRÓSIO, Ana Paula L. **Mineração de Textos**. Goiânia: UFG. 2007. (Série Texto Técnico, INF_005/07).

OLIVEIRA, A. F. **Processamento de linguagem Natural: Princípios básicos e a Implementação de um Analisador Sintático de sentenças da Língua Portuguesa**. Monografia (Ciência da Computação), UFRGS. Porto Alegre – 2013.

OLIVEIRA, N. **Como utilizar as redes sociais?** 19 out. 2011. Disponível em <<http://www.natanaeloliveira.com.br/por-que-utilizar-as-redes-sociais/>>. Acesso em: 13 mai. 2013.

OLIVEIRA, N. **O poder das redes sociais** 12 set. 2012. Disponível em:<<http://www.natanaeloliveira.com.br/o-poder-das-redes-sociais/>>. Acesso em: 17 mai. 2013.

OWEN, Sean et al. **Mahout in Action. Connecticut: Manning Publications Co, 2011. 373p.**

PADAS. 2022. **Biblioteca PANDAS** Disponível em: <<https://pandas.pydata.org/>>. Acesso em: 16 Novembro. 2022.

PLOTLY. 2022. **Biblioteca Requests** Disponível em: <<https://plotly.com/python/plotly-express/>>. Acesso em: 16 Novembro. 2022.

PYTHON. 2022. **Biblioteca Requests** Disponível em: <<https://pypi.org/project/requests/>>. Acesso em: 16 Novembro. 2022.

RENNIE, J. D. et al. **Tackling the poor assumptions of naive bayes text classifiers.** In: ICML. 2003. p. 616-623.

RODRIGUES BARBOSA, Glívia Angélica et al. Characterizing the effectiveness of twitter hashtags to detect and track online population sentiment. In: PROCEEDING OF THE 2012 ACM ANNUAL CONFERENCE EXTENDED ABSTRACTS ON HUMAN FACTORS IN COMPUTING SYSTEMS EXTENDED ABSTRACTS. Austin, 2012.

RUSSEL . Mathew A. **Mining the social web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub and More.** 2 ed. Sebastopol: O'reilly Media, Inc., 2013.

SMAAL , B. **A História do Twitter.** 19 de fev. de 2010. Disponível em: <<http://www.tecmundo.com.br/rede-social/3667-a-historia-do-twitter.htm>>. Acesso em: 20 outubro 2022.

SOUZA , T.; LINDGREN, A. **Mineração de Texto.** Março, 2011. Revista Eletrônica de Informática, vol. 01, n. 1, Março, 2011.

SUMMERFIELD , Mark. **Programação em Python 3: Uma Introdução completa à Linguagem Python.** Rio de Janeiro: Alta Books, 2012.

TAN , Ah-Hwee. **Text mining: The state of the art and the challenges.** In: PROCEEDINGS OF THE PAKDD 1999 WORKSHOP ON KNOWLEDGE DISCOVERY FROM ADVANCED DATABASES, Beijing. 1999.

TWITTER . **TWITTER.** 2013. Disponível em: < <https://dev.twitter.com/>>. Acesso em: 26 ago. 2013.

VOLPATO , B. **Ranking: as redes sociais mais usadas no Brasil e no mundo em 2022, com insights e materiais** 23 Mai. 2022. Disponível em:

<<https://resultadosdigitais.com.br/marketing/redes-sociais-mais-usadas-no-brasil/>>
Acesso em: 20 Outubro. 2022.