

CLASSIFICAÇÃO DE SENTIMENTOS DE REDES SOCIAIS

ALESSON BARRETO DE OLIVEIRA, ELINEY SABINO, ANDREWS ROGER MARTINS PUPO,
IGOR HIDEKI DA SILVA ARAGUSUKU, LEONARDO FRANCA GUIMARAES DE LIMA,
RAMON ALVES TRIGO.

RESUMO:

O intuito deste artigo é catalogar e mensurar mensagens da rede social Twitter, utilizando mineração de dados através do Python e um banco de dados para filtrar todo o conteúdo coletado de acordo com palavras-chave selecionadas e que tem relação com as características de pessoas que apresentam depressão. Durante 4 dias obtivemos a coleta de dados diretamente da rede social citada, referentes a depressão. Essas mensagens foram categorizadas manualmente indicando ou não características da depressão segundo estudos acerca da doença. Com os dados obtidos conseguimos uma visualização gráfica da grande quantidade de dados relacionados ao tema.

Palavras chaves: cyberbullying, redes sociais, mineração de dados.

1. INTRODUÇÃO

A crescente popularização da internet tem ocasionado a maior adesão das pessoas em redes sociais, onde possuem acesso a diversos tipos de conteúdo, que podem incluir até mesmo aqueles que ocasionam malefícios aos usuários, como a depressão por exemplo.

Embora existam leis vigentes e normas de conduta exigidas pelos *sites* acerca dos conteúdos postados e compartilhados, devido a grande quantidade de informações geradas diariamente, não existe um controle eficiente automático sobre os conteúdos em geral. Alguns materiais, como pornografia infantil e imagens chocantes são tiradas do ar rapidamente, pois são reportados pelos próprios usuários. No entanto, conteúdos que podem ocasionar quadros de depressão em pessoas mais sensíveis são bastante subjetivos e, culturalmente mais tolerados na sociedade.

Entre os usuários que apresentam depressão, o sentimento de inferioridade combinado com o *cyberbullying* podem trazer consequências como a tentativa de suicídio e casos ainda mais graves como o suicídio consumado.

Nos últimos anos, técnicas de processamento de linguagem natural (PLN) têm emergido em diversas áreas do conhecimento. Tais técnicas possibilitam o desenvolvimento de ferramentas automáticas

que podem auxiliar no combate dos problemas citados. Através destas ferramentas é possível desenvolver algoritmos capazes de identificar conteúdos agressivos, bem como localizar agredidos e agressores dentro de redes sociais.

Anualmente, aumenta significativamente o número de pessoas com depressão pelo mundo e de acordo com a Organização Mundial de Saúde (OMS), essa doença já é considerada umas das principais causas de problemas e incapacitação de seus portadores pelo globo, onde mais de 300 milhões de pessoas possuem esse transtorno mental mundialmente, revelando sua visível disseminação durante os últimos 12 anos.

Associando esse fato ao crescente número de usuários de redes sociais que criam novas contas todos os anos, esse projeto visa verificar a hipótese de que um *software* para mineração de dados e um banco de dados podem identificar e auxiliar na classificação dos sentimentos dos usuários em uma rede social, que tenham relação com pessoas que possuem sintomas característicos da depressão.

2. REVISÃO DA LITERATURA

A Organização Mundial da Saúde (OMS) afirma que a depressão é o transtorno mais comum no mundo, pois só na última década teve um aumento de 20%. Segundo Viana & Lourenço (2017), existem critérios para a identificação da depressão, que pode ser presenciada por cinco ou mais sintomas como: humor deprimido (em crianças e adolescentes, pode ser humor irritável), acentuada diminuição do interesse ou prazer, insônia ou hipersonia, agitação ou retardo psicomotor, fadiga ou perda de energia, sentimentos de inutilidade ou culpa excessiva ou inapropriada, capacidade diminuída para pensar ou se concentrar, pensamentos recorrentes de morte, ideação suicida recorrente sem um plano específico, uma tentativa de suicídio ou psicologia.

As perturbações depressivas podem ser divididas em duas subcategorias

- a) Perturbação depressiva maior que também é conhecida como depressão unipolar, onde a pessoa tende a ter um humor extremo, no caso o estado depressivo, algumas pessoas acabam apresentando dificuldades no sono, perda de peso e se sentem agitadas e irritáveis. Outras acabam comendo e dormindo em excesso, se sentem dominadas por um sentimento de culpa. Ainda há algumas que podem aparentemente não demonstrar o estado depressivo, exercer bem suas funções no trabalho, em situações sociais, entre outros ambientes.
- b) A distímia é uma forma crônica de depressão, mesmo sendo uma forma menos grave do que a forma mais conhecida da doença, os sintomas podem durar por um longo período de tempo.

Por seus sintomas serem mais leves do que o da depressão comum, acaba dificultando o diagnóstico da doença. A pessoa com distímia acaba apresentando um mau humor e negativismo constante, tem uma irritabilidade excessiva e sentem dificuldade na percepção do prazer.

Segundo Martins (2013), a utilização de redes sociais tem aumentado, principalmente pelos adolescentes, e uma das razões é a possibilidade de se expressar de forma anônima e estar recebendo um *feedback* de um certo grupo de pessoas, porém sua utilização excessiva pode levar o indivíduo ao isolamento de relações face-a-face com a família e amigos.

Um estudo de Valkenburg e Peter (2007) revelou que a comunicação pela Internet teve uma relação positiva quando a proximidade a pessoas já conhecidas foi colocada como uma variável, se tornando um meio de sustentar amizades já existentes. Enquanto a comunicação pela Internet com estranhos teve um efeito negativo para a pessoa, também revelou que pessoas que se comunicam online com amigos e familiares na Internet demonstram um nível menor de depressão, enquanto as pessoas que se comunicam online com estranho demonstraram um nível maior de depressão, um dos motivos é por estarem deixando de investir seu tempo em relações mais próximas com seus familiares e amigos.

Com a crescente utilização de redes sociais, mais pessoas acabam compartilhando suas opiniões, suas experiências de vida e gostos, afirma Kauer (2016). Algumas vezes, essas opiniões acabam sendo representadas na forma de texto não estruturado. Tendo isso em vista, utilizou-se a Análise de Sentimentos (Mineração de Opinião), classificando os sentimentos de acordo com o método que melhor se aplica ao projeto.

3. MATERIAIS E MÉTODOS

Para desenvolver o presente trabalho foi feita a definição da plataforma de desenvolvimento apropriados através da revisão da literatura, estudando técnicas de engenharia de software, principalmente o comportamento dos usuários de redes sociais como o *Twitter*, utilizando mineração de dados e um banco de dados para filtrar todo o conteúdo coletado de acordo com as palavras-chave utilizadas para a execução da mineração.

Durante toda a etapa de mineração de dados, foi utilizado um *script* criado na plataforma de desenvolvimento do Python versão 2.7.14, que se trata de uma linguagem de código aberto (*open source*), além de possuir uma ampla compatibilidade com diversos sistemas operacionais e APIs (*Application Programming Interface*), requisitos de suma importância para o desenvolvimento do projeto.

Para a extração dos dados e sua filtragem, foram utilizados termos como: “depressão”, “triste”, “tristeza”, “suicídio” e “ansiedade”, inseridos dentro do código-fonte para a busca dentro do Twitter, servindo como palavras-chave.

Os termos foram escolhidos, pois de acordo com a revisão da literatura, tem relação direta com as características apresentadas com pessoas que possuem a doença. Além das palavras-chave utilizadas, foram determinadas coordenadas utilizando latitude e longitude através do Google Maps, dos locais onde possuem maiores casos de suicídio, e coincidentemente a depressão, como pode ser visto na Figura 1.

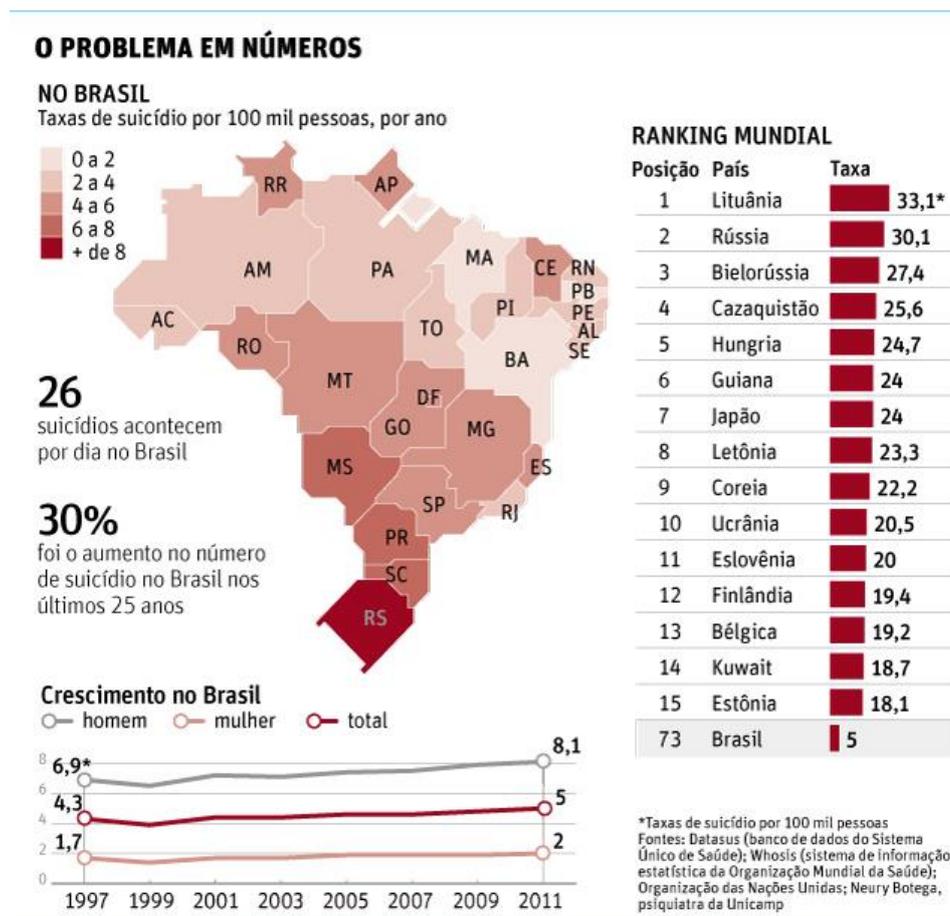


Figura 1 – taxa de suicídio por 100 mil pessoas por ano no Brasil.

Através da Figura 1, foi constatado que o Estado com maior número de suicídios é o Rio Grande do Sul e, por esse motivo, foi utilizado como base para a coleta de *tweets* – que são mensagens que podem possuir até 140 caracteres – dentro do Twitter.

Na Figura 2, está especificado os limites da região dentro do Rio Grande do Sul utilizada para a coleta dos dados via Python utilizando o Google *Maps*.

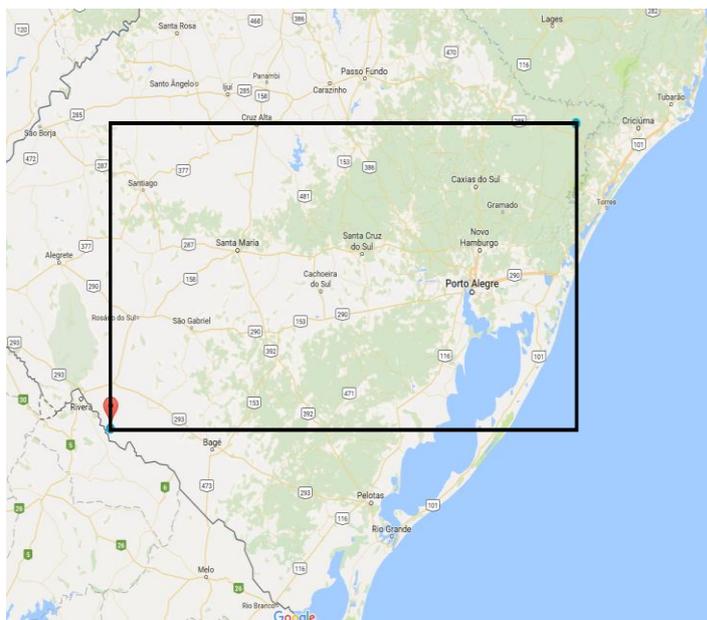


Figura 2 – limites da localidade utilizada como base para a coleta de dados

Fonte: Google Maps

Após a realização da coleta dos dados através do Python, os dados foram visualizados e analisados utilizando um banco de dados chamado *DB Browser for SQLite*, e sua escolha se justifica por ser um programa simples, de fácil manipulação, multiplataforma e *open source*.

4. RESULTADOS E DISCUSSÃO

Durante o dia 31/10/2017 à 03/11/2017, 24hs ao dia, foram coletados 86.120 *tweets* através do *script* utilizado no Python com referências aos termos relacionados à depressão e suicídio. Dos *tweets* coletados, 72.172 se referem ao termo “triste”, 9.230 ao termo “tristeza”, 4.231 a “depressão”, 2.846 a “ansiedade” e 1.912 ao “suicídio”. Os dados foram representados conforme o Gráfico 1.

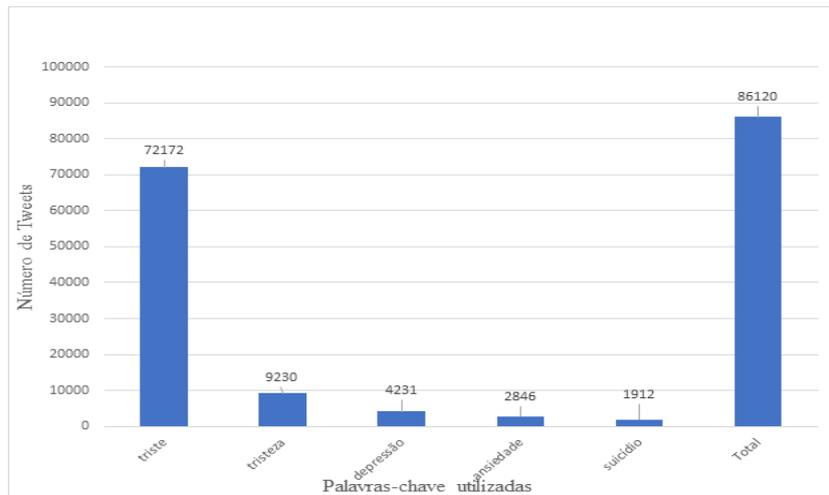


Gráfico 1 - palavras utilizadas relacionadas a depressão

Fonte: Elaborado pelos autores

Apesar de estar evidenciado o grande número de *tweets* das pessoas que possuem características da depressão, em alguns casos os termos foram utilizados de maneira neutra ou com nenhuma associação direta com a doença, como pode ser visto na Tabela 1 preservando o texto original dos *tweets* coletados.

Classificação Manual	<i>Tweet</i> coletado
Possui características de depressão	“depressão é uma coisa tão mais tão séria e ultimamente tenho estado tão perto dessa realidade na família e amizades, é extremamente triste”
Não possui características de depressão	“Seria legal se eles abordassem o tema, muitas pessoas sofrem com a depressão e as pessoas insistem em dizer que é drama, falta do que fazer”

Tabela 1 – classificação manual de depressão de acordo com as características estudadas

Fonte: Elaborada pelos autores

É importante ressaltar que dentro de cada quantidade de *tweets*, existem palavras-chave semelhantes aos demais termos utilizados para a mineração dos dados, influenciando no número final total de cada *tweet* com base no termo utilizado para busca, representados no Gráfico 1.

5. CONSIDERAÇÕES FINAIS

Nesse trabalho foram catalogadas mensagens da rede social Twitter que podem estar relacionadas a pessoas com sintomas de depressão, por meio da mineração de dados. De fato, existem pessoas com características de depressão dentro do Twitter, e a mineração de dados tem o potencial de categorizar essas mensagens, assim como identificar outros tipos de fenômenos dentro do Twitter.

Com os resultados obtidos obteve-se uma visão de como pessoas que postam um *tweet* com mensagens, que tendem a ser de alguém com sintomas de depressão, utilizam as redes sociais como uma forma de expressar o que sentem no momento, evidenciando que é de suma importância haver uma maior atenção para demais técnicas que possam ser aperfeiçoadas, ajudando a identificar não só doenças como a depressão, mas outros fenômenos que podem ser coletados, estudados e identificados dentro das redes sociais.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, R. J. A. Estudo da Ocorrência de Cyberbullying Contra Professores na Rede Social Twitter por Meio de um Algoritmo de Classificação Bayesiano. Instituto Federal de Educação, *Ciência e Tecnologia do Sudeste de Minas Gerais* - Volume 5 – Número 1, 2012.

BORGES, A. P., JOIA, L. A. *Executivos e smartphones: uma relação ambígua e paradoxal*. EBAPE/FGV, 2012.

CARVALHO, A. *Depressão e outras perturbações mentais comuns enquadramento global e nacional e referência de recurso em casos emergentes*. Serviço Nacional de Saúde. Programa Nacional para a Saúde Mental. República Portuguesa. 2017.

KAUER, A. U. *Análise de Sentimentos baseada em Aspectos e Atribuição de Polaridade*. Universidade Federal do Rio Grande do Sul, Instituto de Informática - Programa de Pós-Graduação em Computação, Porto Alegre, 2016.

LOURENÇO, L. M., VIANA, R. S. *Estudo qualitativo sobre a depressão e a ansiedade social na adolescência: uma revisão bibliográfica*. Psicologia.pt - O Portal dos Psicólogos, 2017.

MARTINS, D. A. *Adolescentes internautas, família e depressão: estudo da relação entre a utilização da internet e das redes sociais, o ambiente familiar e a sintomatologia depressiva*. Universidade de Lisboa, Faculdade de Psicologia, 2013.