

INFORMAÇÕES GERAIS DO TRABALHO

Título do Trabalho: Análise de sentimentos no Twitter para a identificação de alunos dos Institutos Federais com Transtorno de Ansiedade.

Autor (es): Nayane Gomes, Carlos Alexandre Silva, Cristiane Targa.

Palavras-chave: Análise de Sentimentos; Ansiedade; Redes Sociais.

Campus: Sabará.

Área do Conhecimento (CNPq): 1.03.03.04-9.

RESUMO

As redes sociais são cada vez mais usadas, seja como forma de entretenimento, comunicação ou liberdade de expressão. Sendo assim, é responsável por gerar grande quantidade de dados, que extraídos podem apresentar um conhecimento útil. Diversos trabalhos têm sido realizados utilizando técnicas de extração de conhecimento através de métodos baseados em aprendizado de máquina, como a análise de sentimentos ou mineração de opinião, sobre textos oriundos de redes sociais, permitindo compreender diversos fenômenos sociais complexos. Um dos problemas sociais que afeta milhares de pessoas em todo o mundo, é a ansiedade, fazendo com que o indivíduo viva com medo do futuro, e não viva o presente. Quando ignorado, pode-se desenvolver uma série de doenças psicossomáticas, ou seja, doenças que afetam a saúde física e mental, como: gastrite, úlceras, colítes, taquicardia, hipertensão, cefaleia, alergias e até mesmo depressão. No âmbito acadêmico, características como 'branco' em prova, dificuldade de prestar atenção e alta distração podem ser indícios de uma mente com níveis significativos de ansiedade. O objetivo deste trabalho é utilizar métodos de análise de sentimentos no Twitter para a identificação de alunos dos Institutos Federais com Transtorno de Ansiedade. Institutos federais são instituições de educação superior, básica e profissional, pluricurricular e multicampi, especializada na oferta de educação profissional e tecnológica nas diferentes modalidades de ensino. Até o presente momento não foram encontrados trabalhos que relacionam transtorno de ansiedade em alunos de institutos federais. Para a identificação dos perfis com transtorno de ansiedade foi necessário a construção de uma base de dados, que foi construída utilizando um algoritmo de rede neural. A rede neural recebe como entrada livros de psicologia e apresenta como saída uma base de dados classificada como ansiosa. Assim, é possível analisar qualquer texto e sentenciar se o mesmo se trata ou não de uma sentença ansiosa. A técnica será aplicada no Twitter de alunos oriundos de 32 Institutos Federais, coletado através de Crawlers, visando identificar usuários com perfil de ansiedade. A criação de uma base de dados exclusiva de ansiedade faz com que pesquisas interdisciplinares na área de psicologia e computação tenham uma referência científica para a produção de novos estudos. Uma vez constatado o problema, é apresentado quais Instituições Federais possuem a maior quantidade de pessoas se expressando em redes sociais, e quais destas são consideradas ansiosas. Fatores como: local, turno, modalidade de estudo, faixa etária ou curso também serão observados. Assim, espera-se que o estudo em questão gere benefícios para o aluno, instituição e pesquisa

INTRODUÇÃO:

A ansiedade tem sido definida como um estado emocional desagradável acompanhado de desconforto somático, que guarda relação com outra emoção - o medo [1]. Esse estado emocional é geralmente relacionado a um evento futuro e, às vezes, considerado desproporcional a uma ameaça real [2]. Pesquisadores têm identificado uma série de cargas crônicas que contribuem para a ansiedade. Dentre elas, ser pobre, ter histórico de minoria racial, estar desempregado, viver em um lar conturbado e ser

solteiro[3]. Para Davidoff, 1980, os transtornos de estudantes universitários giram em torno de problemas acadêmicos e sociais como perda de tempo, sentimento de solidão, e necessidade de corresponder a altos padrões.

As redes sociais têm sido um escape para jovens que não sabem como se expressar no dia a dia. Pessoas com o perfil de ansiedade tendem a se manifestar de forma com que o outro sinta a sua dor, e possa de alguma forma ajudar, mas nem sempre isso acontece, gerando uma frustração muito maior, o que vai só aumentando os níveis de ansiedade e depressão.

Segundo Davidoff [3], participantes de pesquisas sob estado de alta ansiedade exibem uma série de problemas relacionados com a aprendizagem que requeira esforço:

- Pessoas altamente ansiosas têm dificuldade de prestar atenção e distraem-se com facilidade. De fato, o que acontece é que, quando deveriam estar prestando atenção, estão concentradas em irrelevantes como sentimentos de inadequação ou pânico, desempenho dos outros, dor de cabeça, de estômago, e fracasso.
- Os altamente ansiosos usam relativamente poucas das indicações que são fornecidas em tarefas intelectuais. Portanto, ficam menos cientes das implicações e complexidades sendo mais propensos a interpretar erroneamente aquilo que leem, especialmente se o material é difícil ou ambíguo.
- Ao processar informações, os altamente ansiosos não organizam e elaboram apropriadamente, como o fazem as pessoas menos ansiosas.
- Quando a aprendizagem exige flexibilidade, os altamente ansiosos tendem a ser menos adaptáveis em alternar táticas.

Os pesquisadores não compreendem as especificidades, porém sabem que esses aprendizes são incapazes de acessar as informações na memória. Estudantes vítimas de ansiedade em provas e exames declaram que ficam bloqueados e que dá um branco na mente[3].

A motivação para a identificação de alunos com perfis ansiosos dos Institutos Federais se dá ao fato de que, não só a vida pessoal do aluno é afetada, mas sua vida acadêmica também. Indivíduos que estão em situações de desânimo, por exemplo, se veem confrontados com situações que não conseguem ultrapassar, levando o aluno a um estado de apatia e passividade, representado, muitas das vezes por afetos negativos, visto ser difícil aceitar as próprias limitações. [2] Os resultados da análise podem servir também para que as Instituições passem a conhecer a situação do aluno, onde terão um embasamento para possíveis intervenções.

METODOLOGIA:

O processo de identificação de perfis ansiosos consiste em um ciclo de atividades, de forma com que tenha-se um melhor aproveitamento dos dados coletados, com o menor custo computacional, visto que a base de dados requer um alto poder de processamento.

1- Critérios

Ao realizar buscas no twitter foram identificados alguns obstáculos, como: coletar dados apenas de alunos, e, saber se publicação, apesar de citar a sigla, é de fato uma referência à Institutos Federais.

Atributo local de estudo

Diferente de outras redes sociais, a plataforma do Twitter não armazena informações referentes à atividades profissionais, como, escolaridade, local de estudo, ou trabalho. Para contornar este fato, foi adotado a estratégia de armazenar dados de todos aqueles que citaram alguma Instituição Federal em suas postagens.

Foi observado que a chance de analisar um perfil de usuário que não corresponde a um aluno é grande, mesmo citando alguma instituição, por exemplo: *Ana é aluna. Maria não é aluna, mas é amiga de Ana. Ana publica em seu Twitter que haverá um evento no IFMG. Maria visualiza a publicação de Ana, se interessa pelo evento e publica sua postagem. Conclusão: Falso positivo.* Para minimizar o impacto de falso positivo, publicações republicadas foram eliminadas.

Atributo local de postagem

Durante a coleta, foi observado que muitos usuários de outros países postam assuntos que contém a expressão 'IF...', trazendo irregularidades ao estudo em questão. Sendo assim, apenas postagens realizadas no Brasil, cujo idioma é Português, foram aceitas, visto que as outras não se tratavam de estudantes, e sim de pessoas que citavam a mesma palavra, porém com outros significados.

2- Coleta

A extração dos dados foi possível por meio de um Algoritmo que acessa a API do twitter e busca por postagens com alguma referência a Institutos Federais.

Dado uma sigla x, por exemplo 'IFMG', o Algoritmo I, descrito a seguir, percorre o Twitter em busca de postagens relacionadas. Caso encontre, o mesmo começa um processo de verificação das postagens, e só armazena se todos os critérios como idioma, localização e publicação própria forem satisfeitos.

Início Algoritmo I

1. Seja x uma sigla referente a um Instituto Federal;
2. leia (x);
3. **Enquanto** x estiver sendo citado **faça**
4. **se** idioma de x == Português e localização de x == Brasil e x != reepostagem **faça**
5. lista de tweets ← tweet e dados do usuário responsável;
6. **Fim-enquanto**;
7. Planilha recebe lista de tweets;

Fim Algoritmo I

As Publicações que passarem dessas etapas são colocados em uma planilha com atributos, a saber: ID Usuário, Data de publicação, Fonte de publicação e a Publicação.

Após a execução do Algoritmo I, o Algoritmo II é usado para a extração de dados dos usuários identificados como alunos. Dado um id x, por exemplo 'Joao_IFMG', o Algoritmo II percorre o Twitter de x e armazena tudo já publicado, criando-se uma base para cada usuário. As publicações são colocados em uma planilha com os mesmos atributos do algoritmo I.

Início Algoritmo II

8. Seja x um aluno de Instituição Federal;
9. leia (x);
10. **Enquanto** x tiver dados **faça**
11. lista de tweets ← tweet e dados de publicação;
12. **Fim-enquanto**;
13. Planilha recebe todos os tweets de aluno;

Fim Algoritmo II

3- Limpeza e Análise da Base de Dados

A base gerada a partir dos algoritmos do item 2, é pré-processada e analisada. O pré-processamento dos dados é importante, para a remoção de sentenças que não agregam informações relevantes ao texto, tais como, caracteres especiais, termos provenientes da plataforma, e links. Por exemplo, uma linha da tabela gerada é do tipo: *RT @id1: eu não to conseguindo dormir direito, fico acordado pensando no IF <https://t.co/q7gUlhXoxg>* . Após a limpeza, ficaria na forma: *eu não to conseguindo dormir direito, fico acordado pensando no IFMG.*

4- Construção de dicionário

Para garantir o sucesso na detecção dos perfis ansiosos, optou-se pela criação de um dicionário, com uma ampla quantidade de palavras de teor ansioso em português, para que o processo de tradução não interfira na qualidade do texto. As palavras ansiosas recebem o rótulo 1, enquanto palavras denominadas como positivas são identificadas com 0.

5- Rede Neural

A seleção de palavras usadas no dicionário foi criada por meio de uma rede neural. A necessidade de uma inteligência artificial na elaboração de um dicionário se deu por falta de bases atuais existentes. Na literatura, ainda não é possível encontrar uma base de dados classificada como ansiosa.

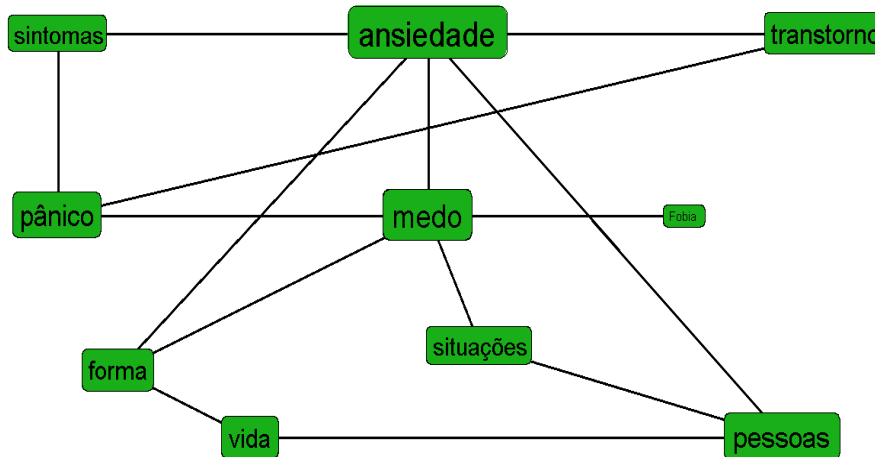
A ferramenta utilizada na construção da rede neural é a biblioteca *Python Gensim*. Essa biblioteca recebe como entrada um conjunto de dados, treina e apresenta como saída, as 10 palavras que mais se relacionam com determinada expressão, como pode ser visto na Figura 1. A base de treinamento usada é composta por dois livros de psicologia com ênfase em Ansiedade, *Mentes Ansiosas* [4] e *Sem Pânico* [5]. Os livros foram escolhidos pelo fato de terem gerado um conteúdo extremamente relevante durante os testes, como demonstra Figura 2, elaborada com o auxílio da ferramenta de mineração *Sobek* [6]:

Figura 1- Palavras relacionadas à ansiedade

Palavras	Porcentagem de aproximação
medo	0.9880714416503906
vida	0.984707236289978
sintomas	0.9830812215805054
tempo	0.981929361820221
corpo	0.9789963960647583
pânico	0.9780239462852478
tratamento	0.9773185849189758
grande	0.9767307043075562
fobia	0.9763853549957275
situações	0.9762625098228455

Fonte: Elaborada pelo autor

Figura 2- Grafo de palavras relacionadas à ansiedade



Fonte: Elaborada pelo autor

6- Análise de Sentimentos

Os *tweets* são analisados através um método próprio, feito na linguagem Python, com a abordagem de aprendizado de máquina. O algoritmo implementado é o de *Naive Bayes*, o mesmo recebe a base ansiosa, cria um modelo de ansiedade, e, faz uma comparação frase a frase para cada linha da base de dados dos alunos. Assim, é possível obter uma definição se determinado *tweet* é ou não ansioso.

RESULTADOS E DISCUSSÕES:

Ainda não é possível descrever o impacto da análise de sentimentos nos Institutos Federais, visto que o presente trabalho se encontra em andamento. Porém, é possível observar que os usuários coletados têm tendência a se expressar de maneira ansiosa, como demonstrado na Figura 3.

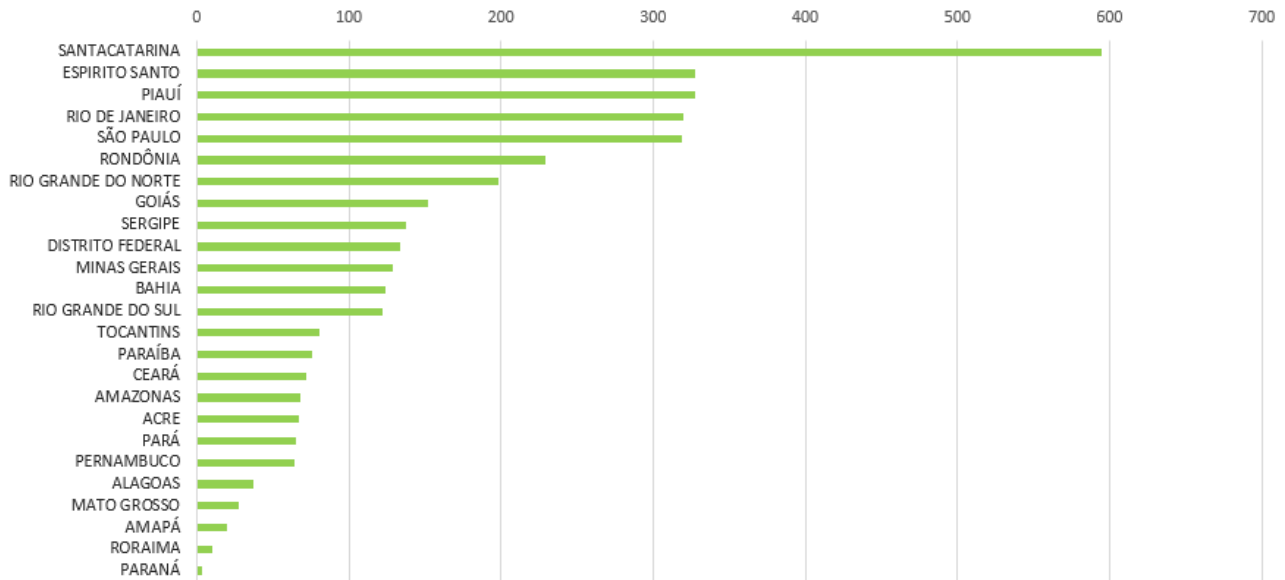
Figura 3 : Publicações de Alunos (ID alterado para preservação de identidade)

ID	Data de publicação	Publicação
48975268	31/07/2018 19:04	Eu to tombada com essa lista 4 de geometria analítica
59726989	31/07/2018 09:51	Eu só queria férias
12332545	30/07/2018 23:49	Estou preocupado de mais 😞😞😞
89578787	30/07/2018 11:00	Por fora parece que eu estou bem Por fora pareço estar legal 🎵
12515141	30/07/2018 10:59	Quando sua ansiedade tá te correndo por dentro mas vc quer se manter positiva e sensata para as pessoas
48484569	30/07/2018 10:59	Tá com saudade de alguém da escola ? — eu quero que todo mundo se f***
30302620	30/07/2018 10:58	Meu deus eu tirei 5 na prova de filosofia to sem acreditar que é real
89572035	30/07/2018 10:58	tenho quinhentos assuntos de matemática p ver hoje e um total de zero vontade mas vamo lá p vida de luta

Fonte: Elaborada pelo autor

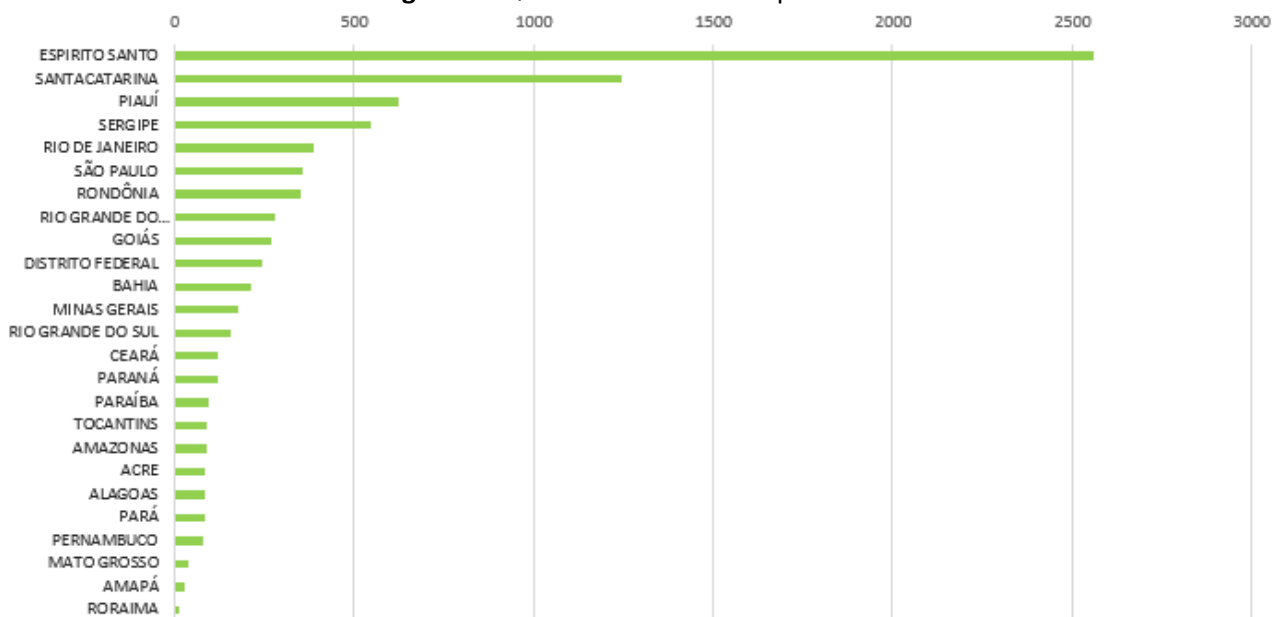
Informações como quantidade de alunos por estado e quantidade de *tweets* por estado também foram observados, como demonstrados na Figura 4 e Figura 5, respectivamente.

Figura 4- Quantidade de alunos por estado.



Fonte: Elaborada pelo autor

Figura 5 - Quantidade de Tweets por estado.



Fonte: Elaborada pelo autor

CONCLUSÕES:

De acordo com o ponto de vista psicológico, a ansiedade é definida como um estado mental carregado de apreensão e recheado de incertezas. Se não controlada, o indivíduo tende a sofrer de outras doenças, físicas ou emocionais. Quando a ansiedade é identificada no âmbito acadêmico, fatores antes estressantes acabam sendo amenizados, visto que as Instituições envolvidas serão notificadas para que entrem com intervenções. A criação de uma base de dados exclusiva de ansiedade produz contribuições significativas na literatura para a produção de novos estudos.

REFERÊNCIAS BIBLIOGRÁFICAS:

- [1] ZAMIGNANI, Denis Roberto; BANACO, Roberto Alves. Um panorama analítico-comportamental sobre os transtornos de ansiedade. Revista Brasileira de Terapia Comportamental e Cognitiva, v. 7, n. 1, p. 77-92, 2005.
- [2] GENTIL, V. Ansiedade e transtornos ansiosos. Valentim Gentil, Francisco Lotufo-Neto e, 1997.
- [3] DAVIDOFF, LINDA L. Introdução à psicologia. 3. ed. São Paulo: Pearson, 2001, p. 98.
- [4] SILVA, Ana Beatriz Barbosa. Mentis Ansiosas. Editora Objetiva, 2011.
- [5] BARRY, Joseph. Sem Pânico.
- [6] REATEGUI, Eliseo et al. Sobek: A text mining tool for educational applications. In: Proceedings of the International Conference on Data Mining (DMIN). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2011. p. 1.