

UMA PROPOSTA DE API BASEADA EM ANÁLISE DE SENTIMENTOS PARA IDENTIFICAÇÃO DE SINTOMAS DE ANSIEDADE NOS INSTITUTOS FEDERAIS VIA REDES SOCIAIS

Lucas Santos Pimenta Junior ¹; Nayane França Gomes ²; Carlos Alexandre Silva ³, Cristiane Norbiato Targa ⁴

1 Lucas Santos Pimenta Junior, Bolsista (CNPq), Bacharelado em Sistemas de Informação, IFMG Campus Sabará, Sabará - MG; lucas.stone.hc@gmail.com;

2 Nayane França Gomes, Bacharelado em Sistemas de Informação, IFMG Campus Sabará, Sabará - MG; nayane.ifmg@gmail.com;

3 Carlos Alexandre Silva, Campus Sabará; carlos.silva@ifmg.edu.br

4 Cristiane Norbiato Targa, Campus Sabará; cristiane.targa@ifmg.edu.br

RESUMO

Diversos trabalhos têm sido realizados utilizando técnicas de extração de conhecimento por meio de métodos baseados em aprendizado de máquina, como a análise de sentimentos, 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. Poucos trabalhos científicos se referem aos aspectos emocionais dos alunos dos Institutos Federais, seja no contexto da psicologia ou no contexto da computação. O presente projeto propõe a elaboração de uma API (Application Programming Interface) baseada em métodos de análise de sentimentos aplicada à rede social Twitter considerando a utilização de Hashtags para a identificação de possíveis alunos de Institutos Federais com sintomas inerentes à ansiedade. Além disso, esse trabalho visa analisar padrões temporais, verificando quais períodos os alunos se encontram mais ansiosos e realizar experimentos utilizando um conjunto de algoritmos de classificação visando obter uma melhor assertividade. De acordo com a revisão bibliográfica realizada pelo proponente do projeto, com exceção de [14], não existe na literatura trabalhos que relacionam análise de sentimentos à identificação de transtorno de ansiedade em possíveis alunos de Institutos Federais. Tendo em vista que a ansiedade representa, segundo especialistas, o pior dos males psicológicos, este projeto visa proporcionar estudos com base psicológica e computacional para analisar e avaliar o quão este sintoma está presente no contexto dos Institutos Federais, podendo futuramente mitigar os efeitos da ansiedade no cenário em estudo. Como contribuição científica, esta proposta pretende impulsionar uma linha de pesquisa no Instituto Federal de Minas Gerais (IFMG) - campus Sabará no desenvolvimento de Machine Learning, bem como promover a pesquisa no campo multi e interdisciplinar, envolvendo especialistas de diversas áreas como da psicologia e estudiosos de linguagens, ambos inseridos no contexto de análise de sentimentos textuais. A contribuição tecnológica é caracterizada pelo desenvolvimento de uma API construída com o conhecimento científico a respeito de Machine Learning e conceitos de psicologia. Esta API poderá auxiliar a equipe psicopedagógica do campus para elaborar instrumentos preventivos à comunidade interna correlacionados com o sintoma de ansiedade. Este fim também caracteriza a contribuição social, pois espera-se que os resultados da análise gerada pelo desenvolvimento do projeto possam servir para que os Institutos Federais vislumbrem a situação dos indivíduos com os quais se relaciona, e que possa haver um retorno de como os alunos percebem os institutos, aprimorando os processos internos do campus, inclusive de avaliação.

INTRODUÇÃO:

As redes sociais são cada vez mais usadas, seja como forma de entretenimento, comunicação ou liberdade de expressão. A cada segundo, vários usuários do Twitter publicam milhões de tweets por dia. Apesar do grande volume de dados produzido, a maioria dos dados gerados em redes sociais são considerados não estruturados, o que torna difícil a sua análise. Existem muitos estudos com o intuito de investigar o conteúdo postado e a maioria está relacionada a análise de sentimentos em redes sociais. A análise de sentimento é baseada no aprendizado de máquina (Machine Learning) e extrai informações subjetivas de textos. Para essa análise é

definida uma base de conhecimento contendo palavras chaves que indicam se existe um significado com relação ao texto em apreciação. A partir desse dicionário são gerados modelos de aprendizado para processar os textos e determinar se seu conteúdo indica uma avaliação sobre o assunto em questão [12]. Devido a sua importância para as empresas e para a sociedade, esse tema não tem ficado restrito apenas a uma área da computação e tem atingido outras áreas. Diversos trabalhos têm sido realizados sobre a análise de sentimentos de textos oriundos de redes sociais. Fornecendo um estudo profundo sobre como emoticons que são utilizados pelos usuários do Twitter, [12] sugere que emoticons podem capturar corretamente sentimentos em mensagens da rede. O UniLex [20] aborda a criação de um método computacional baseado em dicionário léxico, que seja capaz de analisar textos em idioma português brasileiro e de polarizar as palavras de acordo com os sentimentos por elas transmitidos. O estudo em [14] apresenta um modelo que classifica tweets coletados de possíveis alunos de Institutos Federais em ansioso ou feliz. O objetivo do presente projeto é utilizar métodos de análise de sentimentos no Twitter com base em Hashtags para a identificação de possíveis alunos de Institutos Federais com sintomas inerentes à ansiedade. O trabalho apresentado em [14] realiza uma limpeza nos tweets, retirando das postagens as hashtags. Hashtag é uma palavra-chave antecedida pela cerquilha (#), também conhecido como Jogo da velha ou tralha, utilizada para identificar o tema do conteúdo compartilhado nas redes sociais. Cada hashtag criada é transformada em um hiperlink que direciona a pesquisa para todas as pessoas de uma rede social que também marcaram os seus conteúdos com aquela hashtag específica. Ou seja, esta é uma forma prática de agrupar diversos conteúdos sobre um determinado tema, facilitando a pesquisa posterior sobre o tópico desejado. A partir dessa pesquisa é possível encontrar as pessoas que estão comentando sobre o assunto e o que cada uma delas compartilhou. Muitos trabalhos em análise de sentimentos ou retiram dos tweets datas, marcação de outros usuários e hashtags, ou são considerados comentários neutros. Supõe-se que o uso de hashtags pode facilitar a coleta e a identificação de alunos dos Institutos Federais com sintomas inerentes à ansiedade, além de ajudar a identificar pessoas que partilham o mesmo problema.

METODOLOGIA:

Para o desenvolvimento do projeto serão realizadas reuniões semanais ou quinzenais dependendo da atividade a ser realizada conforme o cronograma de trabalho do bolsista. Além disso, a equipe fará uso de Tecnologias da Informação para agilizar a comunicação, como e-mails e WhatsApp. Os softwares necessários para o desenvolvimento da pesquisa são de acesso livre via web, e o material bibliográfico são acessíveis pela internet ou biblioteca do campus. Para o desenvolvimento do presente projeto é necessário um aluno bolsista PIBIC do curso de Bacharelado em Sistemas de Informação do IFMG - campus Sabará com conhecimento em programação. Para alcançar os objetivos do presente projeto, será necessário, nos primeiros três meses de trabalho, realizar um estudo aprofundado da temática de análise de sentimentos, redes sociais e ansiedade. A linguagem de programação sugerida é a linguagem Python, a qual tem disponível inúmeras bibliotecas para estes fins. Durante a implementação, será realizada a coleta de tweets para extrair os dados necessários, mantendo as hashtags e data, bem como a execução de experimentos utilizando outros algoritmos de classificação. Durante a análise de resultados serão verificados padrões temporais, gerando gráficos, identificando períodos onde os alunos se encontram mais ansiosos e a validação da API, junto ao psicólogo, quanto classificação de ansiedade. O bolsista PIBIC é o responsável por todas as etapas de desenvolvimento:

1. Embasamento teórico sobre o problema a ser estudado.
2. Explorar o modelo elaborado por [14].
3. Implementação da API.
4. Análise dos Resultados Gerados.
5. Redação de relatórios/artigos técnico científico.
6. Submissão de resultados relevantes à conferências e/ou periódicos.

RESULTADOS E DISCUSSÕES:

Este trabalho é a continuação do projeto da antiga bolsista Nayane França. O projeto foi renovado e retornou há exatamente três meses, após o retorno temos algumas tarefas realizadas.

1. Coleta de tweets por data temporal.
2. Classificação dos tweets por estação do ano.
3. Leitura do calendário acadêmico para uma comparação entre semanas de provas e nível de ansiedade presente nos tweets.

CONCLUSÕES:

O projeto aqui apresentado dedica-se a estudar aspectos da ansiedade, como identificar possíveis alunos dos institutos federais através de seus posts no Twitter utilizando métodos de análise de sentimentos. Sendo assim, alinham-se entre os resultados esperados desta investigação, os seguintes aspectos:

- Coleta de tweets de prováveis alunos. Indicadores quantitativos: Número de Institutos Federais identificados; Número de Tweets Coletados;
- identificação do melhor algoritmo de classificação para o problema proposto. Indicadores quantitativos: Quantidade de acertos na classificação dos tweets utilizando a Matriz de confusão. Além disso, existem alguns parâmetros que precisam ser observados para mensurar a eficiência:

1. Acurácia: Percentual de acertos.
2. Precisão: Número de vezes que aconteceu a predição correta de uma classe, dividido pelo número de predições da classe.
3. Revocação: Número de vezes que aconteceu a predição correta de uma classe, dividido pelo número de vezes que a classe aparece no teste.
4. Medida F1: Média entre precisão e revocação.

A implementação da API, objetivo final desse projeto, poderá auxiliar a equipe psicopedagógica do campus a elaborar instrumentos preventivos inerentes à ansiedade. Este fim também caracteriza a contribuição social, pois espera-se que os resultados da análise possam servir para que os Institutos Federais passem a conhecer a situação dos indivíduos com os quais se relacionam, onde terão um embasamento para possíveis intervenções.

REFERÊNCIAS BIBLIOGRÁFICAS:

[12] BENEVENUTO, Fabrício; RIBEIRO, Filipe; ARAÚJO, Matheus. Métodos para análise de sentimentos em mídias sociais. 2015.

[14] GOMES, N. F.; SILVA, C. A.; TARGA, C. N. Análise de sentimentos no Twitter para a identificação de alunos dos Institutos Federais com Transtorno de Ansiedade. In. VII Seminário de Iniciação Científica IFMG - SABARÁ, vol. 1, pp. 309-316, Sabará, MG, 2018.

[20] YAGUI, M. M. M., MAIA, L. F. M. P., UGULINO, W., VIVACQUA, A., and OLIVEIRA, J. "Bela, Recatada e do Lar": Base de Dados e Aspectos do Movimento Social Ocorrido na Rede Social Online Twitter. In Proceedings of the XXXVII Congress of the Brazilian Computer Society: #ComputaçãoParaTudoeParaTod*s, pages 1585– 1594, São Paulo, 2017.