

Método para Contrastar Percepção de Suporte Social com Dados Extraídos do *Twitter* em Português

Cássio de Alcantara¹, Elaine Ribeiro de Faria Paiva¹,
Maria Camila Nardini Barioni¹, Luiz Carlos Oliveira Junior¹

¹Instituto de Computação – Universidade Federal de Uberlândia (UFU)
Caixa Postal 38408-100 – Uberlândia – Minas Gerais – Brazil

alcantara.cassio@ufu.br, elaine@ufu.br, camila.barioni@ufu.br,

luizcarlosojr@ufu.br

Abstract. *The study of characteristics such as behavior, temperament, and social support of individuals can be used to anticipate psychological disorders. In particular, online social networks can be used as a source of information to understand such characteristics. The feelings present in a user's posts, along with their behavior on an online social network, can be used to understand their perception of social support. This study proposes a method for contrasting a user's behavior on the online social network Twitter with their perception of social support, measured through a questionnaire. To do this, firstly, an unsupervised model was proposed for classifying the polarity of posts, which obtained superior results in all three tested datasets when compared to the Vader and SentiStrength models, reaching an F1 score of 57% in the best case. To contrast the questionnaire results with the user's behavior on Twitter, Pearson and Point-Bisserial correlation coefficients were calculated, which provided useful information to understand this relationship, complementing the information obtained through questionnaires.*

Resumo. *O estudo de características como comportamento, temperamento e suporte social de pessoas pode ser usado para antecipar transtornos psicológicos. Em especial, as redes sociais online podem ser usadas como fonte de informação para entender tais características. Os sentimentos presentes nas postagens de um usuário em conjunto com o seu comportamento em uma rede social online podem ser usadas para entender sua percepção de suporte social. Este estudo propõe um método para contrastar o comportamento de um usuário da rede social online Twitter com a sua percepção de suporte social, medida por meio de um questionário. Para isso, primeiramente, foi proposto um modelo não supervisionado para classificação da polaridade das postagens que obteve resultados superiores em todas as três bases testadas quando comparado com os modelos Vader e SentiStrength, atingindo F1 de 57% no melhor caso. Para contrastar os resultados do questionário com o comportamento do usuário no Twitter foram calculados os coeficientes de correlação de Pearson e Ponto-Bisserial, que forneceram informações úteis para entender tal relação, complementando as informações obtidas por meio de questionários.*

1. Introdução

A análise de redes sociais se mostra útil em diversos cenários. Na área de propaganda e *marketing* pode ser utilizada para entender melhor o consumidor, auxiliando, por exemplo, na propagação de marcas. No campo da epidemiologia, pode ser usada para entender propagações de doenças através de interações sociais [Mikolajczyk and Kretzschmar 2008]. Além disso, as possibilidades só aumentam com o surgimento e crescimento das redes sociais *online*, onde é evidenciada a característica de formar grupos da sociedade humana [Castells and Blackwell 1998].

Com o crescimento do uso das redes sociais *online*, acredita-se ser relevante construir representações para descrever o comportamento das pessoas, podendo contribuir na área da saúde com a identificação antecipada de distúrbios psicológicos. Apesar de existirem alguns estudos descritivos no sentido de entender o comportamento dos usuários na rede social *online* [Lima and de Castro 2019], esses são ainda escassos.

Na psiquiatria, a Escala de Satisfação com o Suporte Social (ESSS) é uma medida que pode ser relacionada com problemas de saúde física, condições de trabalho, estresse [Hohaus and Berah 1996], dentre outros aspectos. Ainda que a percepção de suporte social de um usuário seja tema de estudo na psiquiatria, a relação dessa medida com o comportamento de um usuário na rede social *online*, ainda não foi explorada na literatura correlata sobre análise de dados de redes sociais.

Uma das formas de se entender o comportamento de um usuário na rede social *online* é entendendo o sentimento (positivo, negativo e neutro) expresso nas suas postagens. Várias abordagens tem sido propostas na literatura para análise de sentimentos em diferentes contextos [Burth et al. 2019]. A maioria desses trabalhos usam abordagens supervisionadas, que exigem uma base de dados rotulada, as quais são difíceis de serem obtidas. Por conta disso, boa parte dos trabalhos reutilizam bases de dados públicas ou aplicam técnicas de *self-report* na etapa de coleta [Cohan et al. 2018]. Em consequência disso, acredita-se que os estudos possam ficar limitados, pois podem acabar sendo reféns do contexto da base utilizada, mas também que existe oportunidade de melhoria na qualidade das pesquisas ao se propor soluções que contornem esse cenário.

Abordagens não-supervisionadas também são exploradas por autores e apresentam bons resultados no geral [Freitas 2013]. Entretanto, existe uma carência de bons métodos produzidos para a língua portuguesa, o que leva autores a utilizarem métodos como o Vader [Thelwall et al. 2010] com bases traduzidas para o inglês ou até mesmo traduzindo o léxico [Maia et al. 2021].

Diante do cenário exposto, este trabalho propõe um método para contrastar dados da rede social *online* *Twitter* com a percepção de suporte social, medida pelo questionário ESSS. Para isso, um usuário é representado pelo sentimento expresso nas suas postagens e por um conjunto de atributos comportamentais do mesmo. Para a classificação do sentimento das postagens, este trabalho propõe uma nova abordagem não-supervisionada baseada em léxicos, evitando assim a criação de bases de dados rotuladas.

O restante do trabalho se estrutura na seguinte maneira: A seção 2 apresenta trabalhos relacionados ao presente estudo, a seção 3 traz os principais conceitos que serão utilizados no experimento, na seção 4 tem-se a proposta metodológica, na 5 são mostrados os experimentos e seus resultados e, por fim, a seção 6 finaliza com as considerações

finais do trabalho e próximos passos.

2. Trabalhos Relacionados

Existem diversos trabalhos cujo objetivo é estudar o comportamento dos usuários em redes sociais *online*. Para tal, existem vários objetos de estudo, como as publicações (ou postagens), os perfis dos usuários ou até mesmo o comportamento e evolução de um determinado assunto através de *hashtags* ou palavras-chave [Paes et al. 2022]. Dados oriundos do *Twitter* são amplamente utilizados para realizar estudos de variados temas, sendo a maioria envolvendo a análise dos sentimentos expressados nas postagens, principalmente classificando-as em positivo, negativo e neutro. Tais sentimentos por muitas vezes são relacionados a um evento ou um tema em específico de análise, como atentados terroristas, eleições, esportes e muito mais [Hott et al. 2022].

Na área médica, em específico na psicologia, alguns autores realizaram estudos buscando relacionar testes de personalidade com redes sociais, como [Lima and de Castro 2019], que buscaram descrever as características comportamentais esperadas de cada temperamento resultante do teste de Myers-Briggs. Na mesma linha, [Cohan et al. 2018] abordaram técnicas para identificar usuários que possuem onze diferentes problemas de saúde mental (ex. Autismo, Bipolaridade, etc) utilizando os textos de suas postagens no site *Reddit*.

Além desses trabalhos, a utilização de modelos supervisionados aparece em diversos trabalhos de Processamento de Linguagem Natural (PLN) envolvendo saúde mental. Apesar de apresentarem bons resultados, a dependência de bases rotuladas é um problema comum que os autores enfrentam [Fonseca et al. 2021]. Para contornar, estratégias, como coletar bases utilizando técnicas de *self-report* [Plank and Hovy 2015] ou a criação de comitês para rotulação de base [Burnap et al. 2017], são alternativas encontradas. Mas essas abordagens podem trazer limitações, como a dependência de uma boa rotulação, ou a necessidade de possível revisão das tabulações para garantia de qualidade.

Abordagens não supervisionadas também são exploradas e apresentam bons resultados, como [Lima and de Castro 2019], que em seu trabalho utilizaram léxicos para gerar variáveis referentes às publicações somadas com variáveis comportamentais do usuário (ex.: quantidade de postagens.) a partir de sua base de dados. Já [Santos and Goya 2022] criaram um método não supervisionado para rotulação automática do posicionamento de um usuário do *Twitter* em relação a um tema político polarizado que foi capaz de rotular 98% da base estudada em seu trabalho.

A criação de léxicos, apesar de ser normalmente evitada, mostra na literatura que oferece os resultados confiáveis, como constata [Freitas 2013]. Nesse caminho, os trabalhos de [Thelwall et al. 2010] e [Hutto and Gilbert 2014] propuseram a criação de modelos baseados em léxicos para resolver problemas de PLN que são muito usados na literatura atual, mas foram desenvolvidos para a língua inglesa, podendo não performar bem quando utilizados em bases em português.

Apesar de todas as contribuições citadas, constatou-se carência de pesquisas em português que contribuam para a área médica ao constatar o comportamento *online* com o real. Especificamente, até o momento desta pesquisa, não foram encontrados trabalhos que abordem percepção de suporte social nesse contexto. Sendo assim, o presente traba-

lho visa contribuir para a área da saúde, provendo meios que auxiliem na avaliação além dos questionários comumente aplicados para extrair conhecimento.

3. Conceitos Fundamentais

O Questionário de Escala de Satisfação com o Suporte Social¹ (ESSS) [Ribeiro 1999] consiste, em sua versão final, em 15 afirmações das quais o respondente deve assinalar o grau de concordância em uma escala Likert com cinco posições, indo de 1 a 5. Essas afirmações são divididas em quatro dimensões e o resultado final (esss) do questionário corresponde a soma de todas as respostas. As quatro dimensões consideradas são descritas abaixo:

- **Satisfação Amigos:** Resultado da soma das respostas das primeiras cinco perguntas do questionário, pertencendo ao intervalo [5, 25].
- **Intimidade:** Resultado da soma das respostas das perguntas 6, 7, 8 e 9 do questionário, pertencendo ao intervalo [4, 20].
- **Satisfação Família:** Resultado da soma das respostas das perguntas 10, 11 e 12 do questionário, pertencendo ao intervalo [3, 15].
- **Atividades Sociais:** Resultado da soma das respostas das perguntas 13, 14 e 15 do questionário, pertencendo ao intervalo [3, 15].

A partir dos valores de cada dimensão, assim como o ESSS, é possível conduzir diversas análises e comparações com outros indicadores, como a escala de estresse de indivíduos [Hohaus and Berah 1996].

4. Método

Para atingir o objetivo de comparar a percepção de suporte social com o comportamento do usuário no *Twitter*, o trabalho apresentado aqui apresenta um método que é dividido em duas etapas conforme ilustrado na Figura 1. Na primeira etapa o foco é a construção de um método não supervisionado para classificação de polaridade de textos enquanto, na segunda etapa, a finalidade é contrastar o comportamento de usuários do *Twitter* com as suas respectivas respostas do questionário ESSS. As seções a seguir detalham cada uma dessas etapas.

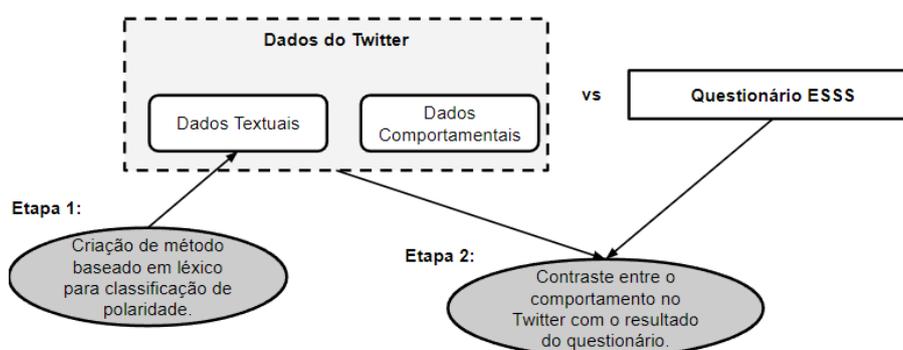


Figura 1. Representação gráfica do método proposto

¹O questionário pode ser visualizado no seguinte link: <https://forms.gle/WvvyzCAv5xZpaDs9>

4.1. Criação de Método para Classificação de Polaridade Baseado em Léxico

O método não supervisionado para classificação de polaridade baseada em léxico e regras de linguagem desenvolvido no trabalho apresentado aqui foi denominado *Lexical Portuguese Approach for Polarity Classification* (LeXPAPC). Ele segue o fluxo mostrado na Figura 2. Cada passo do método está descrito a seguir:

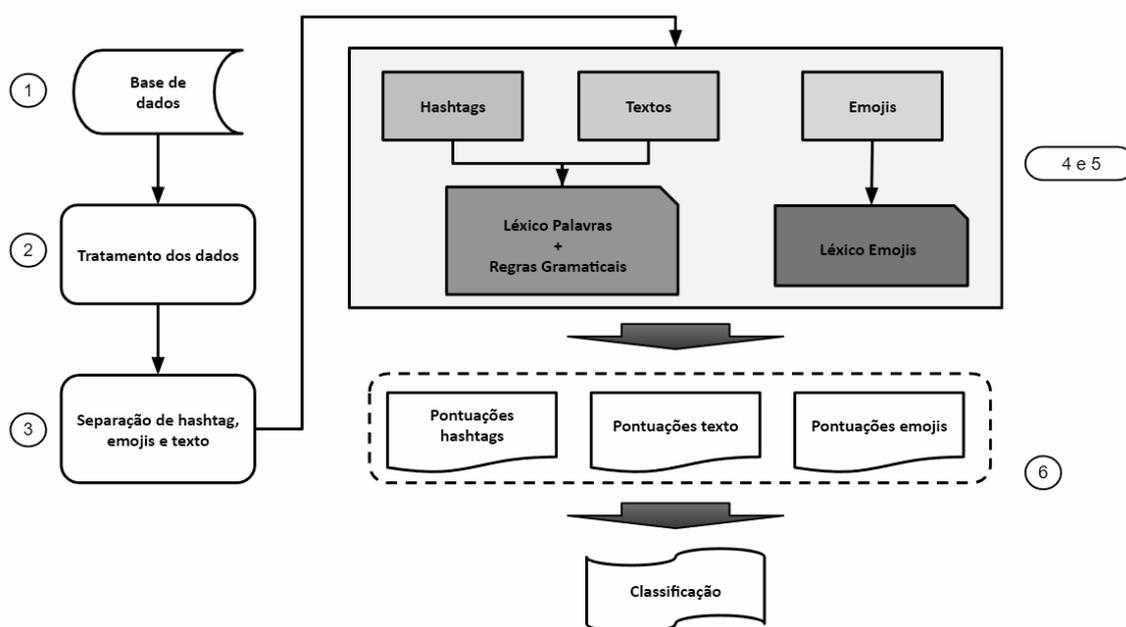


Figura 2. Representação gráfica do método LeXPAPC.

1. **Base de dados:** O método foi desenvolvido para aceitar qualquer base de dados textual, principalmente com características de redes sociais *online*, que contém *emojis*, *links*, citações de usuários, compartilhamento de arquivos, etc;
2. **Tratamento dos dados:** Esta etapa consiste na limpeza dos registros, para tal todos os caracteres são colocados em caixa baixa, são removidos acentos, pontuação, links, citação de usuários, imagens, *Gifs*, vídeos e palavras de parada (*stopwords*);
3. **Separação de Vetores de Entidades:** Cada registro é separado em três vetores chamados de entidades, são elas: **hashtags**, **textos** e **emojis**. Cada um é preenchido com o seu tipo de dado correspondente (ex.: o vetor de entidades de **emojis** é um vetor contendo apenas **emojis**). Um registro pode retornar um ou mais vetores de entidade vazios. Caso os três sejam vazios, esse registro não passará para as próximas etapas e não será classificado;
4. **Consulta nos Léxicos:** Para cada elemento dos vetores ou combinação de até três elementos, e em cada vetor entidade da etapa anterior, é atribuído um valor caso ele exista em um dos léxicos. Textos e *hashtags* são consultados no léxico de palavras e *emojis* são consultados no léxico de *emojis*. Ambos léxicos são constituídos de duas colunas, sendo a primeira a chave e a segunda o peso, que pode ser -1 ou 1 . A criação dos léxicos será abordada em 4.1.1;
5. **Regras Gramaticais:** Assim como no Vader, entende-se que termos que expresam adversão podem ser tratados de modo separado [Hutto and Gilbert 2014].

Para tal, caso os termos "mas", "entretanto", "sqn" ou "porem" apareçam em algum vetor, os sinais dos valores dos pesos anteriores são invertidos;

6. **Classificação Final:** Cada vetor de entidade retornará um vetor de pesos e eles são somados, resultando em um valor para cada um. Os valores resultantes são utilizados para calcular a classificação final que assume o valor final da soma de um dos vetores de entidade, obedecendo à hierarquia *emoji*, *hashtag* e, por último, texto define o resultado. Caso o valor seja positivo a classificação fica como "positivo", negativo fica como "negativo" e apenas vetores que somam 0 que não sejam vetores de zeros resultam na classificação "neutro", os demais são considerados "não conclusivos". Para diminuir a incidência dessa última classificação, duas possibilidades de classificação também foram testadas a partir dos resultados da classificação dos algoritmos Vader e SS nas bases:

- **Funil:** Considera-se como classificação final o primeiro modelo que obter classificação que não seja "não conclusivo", obedecendo a seguinte hierarquia: método LeXPAPC, Vader e SS;
- **Votação:** Considera-se como classificação final o voto da maioria dos três modelos: LeXPAPC, Vader e SS.

Na Figura 3, tem-se um exemplo de como um registro se comporta ao ser submetido ao método descrito. Na segunda etapa o texto foi colocado em caixa baixa, foram eliminadas as pontuações, a citação "@luiz397" e as palavras de parada "estou", "o", "do" e "que". Na terceira etapa foram gerados três vetores a partir do vetor resultante da segunda etapa, sendo que o vetor de *hashtags* ficou vazio. A quarta gerou novos vetores contendo os pesos das palavras e emojis após consulta nos respectivos léxicos, sendo que a palavra "mas" foi mantida para a próxima etapa, na qual os valores anteriores a ela foram invertidos. A sexta etapa somou os pesos de cada vetor, chegando em dois valores, e a classificação final obedeceu à hierarquia, ficando com o valor dos emojis e resultando "neutro".

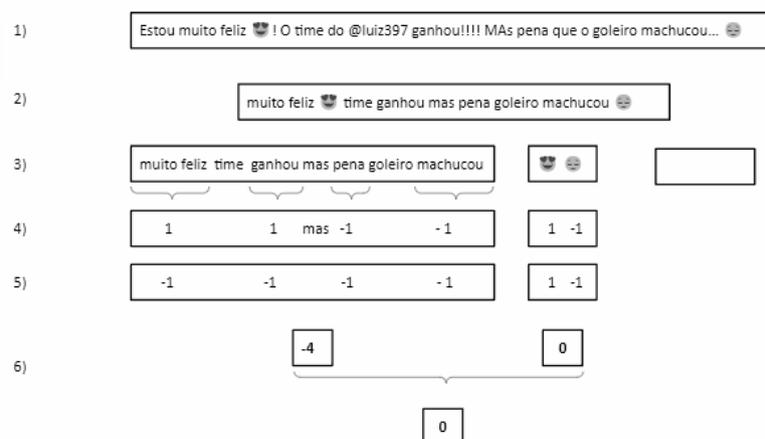


Figura 3. Exemplo do comportamento de um registro submetido ao LeXPAPC

4.1.1. Criação dos Léxicos

Para a criação dos léxicos, primeiramente, fez-se uma análise nos algoritmos do Vader e SS, onde é possível notar que a característica de atribuir um peso que reflete a intensidade de todas as palavras contidas nos registros podem acarretar alguns problemas, como, por exemplo, ao submeter a frase "eu saí da miséria" ao Vader, ela receberá uma classificação final negativa (-0,5719) por conta do peso de -1,7 da palavra miséria, que pode ser usada em contextos positivos e negativos, mas que, provavelmente, na base de dados utilizada para criar o método era um contexto negativo. Além disso, existem situações em que o peso das palavras faz com que o resultado fique errado, como a frase "o caos provoca a minha alegria" que recebe classificação final negativa pelo Vader (-0,2732), isso porque as palavras possuem pesos que não garantem a classificação correta.

Na tentativa de contornar o problema identificado anteriormente, os léxicos foram construídos isolando os itens (palavras e *emojis*) e atribuindo pesos, independente do contexto das bases de dados. Além disso, apenas palavras e expressões de no máximo três palavras (ex.: chutar cachorro morto.) que apareceram com frequência acima de 5% receberam pesos, sendo apenas 1 ou -1. O processo de atribuição de pesos foi realizado em rodadas com as bases de treino, que foram todas somadas em uma única base, sendo que, em cada rodada, novos itens foram adicionados nos léxicos e testados com o método descrito em 4.1. Só ficaram nos léxicos resultantes aqueles itens que apresentaram alguma melhora nas medidas de avaliação.

4.1.2. Avaliação do LeXPAPC

Para avaliar o LeXPAPC, as bases foram separadas em treino e teste na proporção de 70% e 30%, respectivamente. A base de treino foi utilizada para criar o método, enquanto a de teste para avaliá-lo utilizando as medidas acurácia, precisão, f1 e abrangência. A avaliação foi realizada comparando os resultados do LeXPAPC com os métodos Vader e SentiStrength (SS) nas mesmas bases.

4.2. Contraste entre o Comportamento no *Twitter* e ESSS

Para contrastar o comportamento na rede social *online Twitter* com os resultados do questionário ESSS são calculados os valores dos **Coefficiente de Correlação de Pearson** para pares de variáveis contínuas e o **Coefficiente de Correlação de Ponto-Bisserial** para uma variável dicotômica e uma contínua. Esses foram escolhidos por se comportarem bem em situações semelhantes ao problema tratado neste artigo, sendo amplamente utilizados na literatura. As análises se dão por meio da observação do comportamento desses valores.

As variáveis utilizadas na comparação são criadas a partir da saída da API do *Twitter* e separadas em dois tipos: **variáveis comportamentais** (x1, x2, x3 e x6) e **variáveis textuais** (x4 e x5). Essas foram somadas às variáveis resultantes do questionário ESSS (y1, y2, y3, y4 e y5). A descrição de cada variável e como foi obtida está listada a seguir:

- **satisfacao-amigos (y1)**: Calculado conforme descrito na seção 3;
- **intimidade (y2)**: Calculado conforme descrito na seção 3;
- **satisfacao-familia (y3)**: Calculado conforme descrito na seção 3;
- **atividades-sociais (y4)**: Calculado conforme descrito na seção 3;

- **esss (y5)**: Calculada somando os valores das variáveis listadas acima, pertencendo ao intervalo [15, 75];
- **qtd-seguidores (x1)**: Quantidade de seguidores que o usuário possui no momento da coleta, vindo direto da API;
- **qtd-amigos (x2)**: Quantidade de usuários que o usuário segue no momento da coleta, vindo direto da API;
- **genero (x3)**: Autodeclarado pelos usuários nas respostas do questionário aplicado, sendo 1 para masculino e 0 para feminino;
- **media-curtidas (x4)**: Média de curtidas por postagem de cada usuário. Calculada a partir da quantidade de curtidas em cada postagem vinda da API;
- **perc-postagens-positivas (x5)**: Após a aplicação do método criado e validado na etapa anterior foi calculada a porcentagem de postagens positivas para cada usuário;
- **perc-atividade-noturna (x6)**: A partir do campo *created_at* da API, foi extraída a hora da postagem e separada em turnos diurno e noturno, sendo considerado a partir das 18hrs como noturno. Após, foi calculada a porcentagem de noturno;
- **perc-uso-media (x7)**: A partir do campo *image_link* da API, foi criada uma variável binária indicando a presença de imagens nas postagens. Posteriormente foi calculada a porcentagem da mesma.

5. Experimento e Resultados

Nessa seção será descrito o experimento realizado separado em duas etapas, sendo a primeira referente ao LeXPAPC e a segunda referente ao contraste entre o comportamento no *Twitter* e os resultados do questionário ESSS. Em cada etapa também serão descritas as bases de dados utilizadas, que são todas em português e coletadas do *Twitter* e do *Instagram*, sendo que na segunda etapa apenas base do *Twitter* foi utilizada.

5.1. Etapa 1: Criação do LeXPAPC

Para esta etapa foram utilizadas bases de dados, com cada registro classificado em uma das três possíveis classes: positivo, neutro ou negativo. A quantidade de registros de cada base original não coincide com a quantidade utilizada nesse trabalho por conta das políticas de cada *site* (ex.: Perfil bloqueado na rede social.). A seguir serão apresentadas cada uma, suas características e origem:

- **Tash-pt** - É uma base de dados, com 2.787 registros (1.520 após a coleta), referente a rede social *online Twitter* e coletada pela API do *site*. A coleta foi por palavras-chave, extraídas do TeP 2.0 ². A rotulação foi realizada através de uma interface *web* desenvolvida pelos autores desse trabalho, onde cada postagem foi rotulada por até cinco pessoas e a classificação final foi determinada pela maioria dos votos.
- **TweetSentBR** - A TweetSentBR é uma base de dados referente ao *Twitter* e possui 15.000 registros (9.832 após a coleta), sua coleta foi realizada por *hashtags* de programas de televisão, ignorando as postagens de figuras públicas, empresas, etc. A rotulação foi computada pelo voto da maioria e feita por sete profissionais das áreas de jornalismo, ciências da computação e letras.

²<http://www.nilc.icmc.usp.br/tep2/index.htm>

- **Base TEMPS-RIO** - A Base TEMPS-RIO [Martins et al. 2022] foi coletada, com a aprovação do Comitê de Ética em Pesquisas com Seres Humanos — CAAE 35533420.2.0000.5152, que contém 495 registros referentes a rede social *online Instagram*. A rotulação foi realizada por três alunos de graduação seguindo um manual de orientações desenvolvido pelo grupo de pesquisa e a classificação final foi determinada pela maioria.

O LeXPAPC foi avaliado em cada base de teste, assim como os métodos Vader e SS. Os resultados estão dispostos na Tabela 1. Analisando-a é possível verificar que a acurácia e o f1 do LeXPAPC se sobressai nas três bases comparando com Vader e SS, mas, em contrapartida, observando a Figura 4, notamos que a abrangência do modelo não é satisfatória para as bases Tash-pt e TweetSentBR. Em contrapartida, Votação e Funil apresentaram bons resultados e a abrangência ficou em 100% para todas as bases, sendo que Funil ficou levemente melhor, portanto, é a melhor estratégia dentre as testadas.

Base	Método	Acurácia (%)	Precisão (%)	F1 (%)	Abrangência (%)
Tash-pt	Vader	0,4529	0,4615	0,4646	1
	SS	0,4385	0,4323	0,4316	1
	LeXPAPC	0,5398	0,4331	0,4698	0,31
	Votação	0,4713	0,4741	0,4723	1
	Funil	0,4700	0,4718	0,4709	1
TweetSentBR	Vader	0,5913	0,5542	0,5626	1
	SS	0,4737	0,4942	0,4850	1
	LeXPAPC	0,7770	0,5586	0,5526	0,36
	Votação	0,5675	0,5792	0,5710	1
	Funil	0,5726	0,5799	0,5717	1
TEMPS-RIO	Vader	0,5293	0,4770	0,4933	1
	SS	0,5071	0,4882	0,5045	1
	LeXPAPC	0,5347	0,4278	0,5410	0,27
	Votação	0,5556	0,5173	0,5090	1
	Funil	0,5697	0,5360	0,5528	1

Tabela 1. Métricas de todos os métodos nas 3 bases

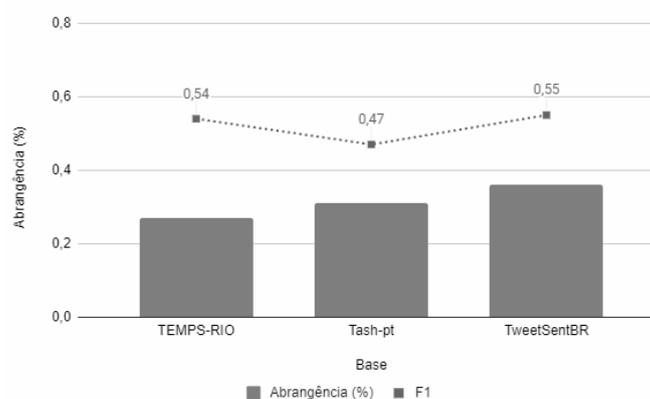


Figura 4. Abrangência e F1 do LeXPAPC em cada base

5.2. Etapa 2: Contraste entre ESSS e o comportamento no *Twitter*

Para esta etapa do experimento foi utilizada uma base chamada de **Base ESSS** que também foi obtida na coleta da **Base TEMPS-RIO** e seguindo a mesma aprovação do comitê de ética. Essa foi coletada da rede social *online Twitter* através da API oficial ³, contendo 2.894 registros referentes a 18 usuários e não foi rotulada com polaridade. Para cada usuário também tem-se as respostas do questionário ESSS descrito na seção 3.

Na Tabela 2 têm-se os coeficientes de correlações para cada par de variável e, na Figura 5 tem-se um exemplo visual da relação entre o par de variáveis **perc-atividade-noturna** e **esss**. A seguir têm-se algumas afirmações que podem ser concluídas após observar os resultados:

- Quantidade de amigos só contribui positivamente para intimidade;
- Atividade noturna é a variável mais relevante no geral e contribui positivamente em todos os casos, atingindo o maior valor de correlação com **esss**;
- Uso de imagens nas postagens é a variável menos relevante;
- O maior valor de gênero é para satisfação família e ele é negativo. Nesse caso, quer dizer que mulheres apresentam valores maiores de satisfação família;
- Ainda sobre o gênero, homens apresentam valores superiores apenas em intimidade;
- Quanto menos seguidores, maiores os valores de satisfação amigos e intimidade.

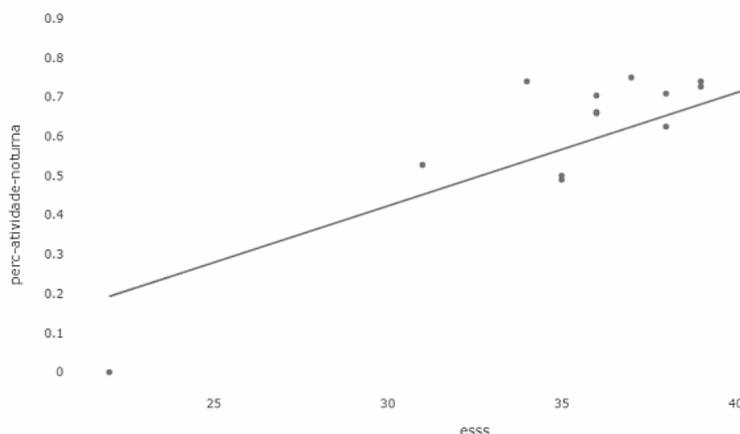


Figura 5. Relação entre atividade noturna e ESSS

	x1	x2	x3	x4	x5	x6	x7
y1	-0,2338	-0,5018	-0,1824	0,3975	0,3291	0,4653	0,2757
y2	-0,2297	0,1243	0,2039	0,1463	-0,3748	0,1087	0,1314
y3	0,0965	-0,4215	-0,5291	-0,1513	0,3970	0,5821	-0,1100
y4	0,3732	-0,0517	-0,1534	0,3287	0,4886	0,4361	0,2075
y5	0,0379	-0,3988	-0,3104	0,3851	0,4428	0,7735	0,2666

Tabela 2. Correlações de Pearson e Ponto-Bisserial

³<https://developer.twitter.com/>

6. Conclusões

As principais contribuições do trabalho descrito aqui compreendem a proposta de um método não supervisionado baseado em léxico para classificação de polaridade em postagens da rede social *online Twitter* (LeXPAPC) e a realização de uma análise contrastando os resultados do questionário ESSS com o comportamento do usuário na rede.

Sobre o LeXPAPC, foi visto que, comparando o método proposto com os demais, os resultados foram superiores, apesar de a abrangência ter sido inferior, dado que os métodos comparados classificam todas as entradas. Contudo, ao aplicar estratégias de combinação de métodos, em especial o funil, houve uma queda na acurácia, mas as demais medidas se tornaram mais consistentes.

A análise de contrastes mostrou relações entre o comportamento no *Twitter* e o questionário que podem auxiliar a área da saúde, possibilitando agregar a realidade *online* nas análises e até mesmo poder agir preditivamente com devido acompanhamento.

Para trabalhos futuros pretende-se, para o LeXPAPC, investir na tentativa de adicionar níveis de pesos nas palavras, diferenciando-as entre si assim como os demais métodos expostos neste trabalho fazem. Também será estudada a inclusão de técnicas para identificação de intensidade na escrita por pontuação e/ou letras repetidas em palavras para decidir possíveis empates entre os pesos. Na análise de contraste pretende-se ajustar modelos preditivos a partir das variáveis apontadas como relevantes nesse trabalho, além disso, os resultados serão reaproveitados para análises similares na base **Base TEMPS-RIO**.

Referências

- Burnap, P., Colombo, G., Amery, R., Hodorog, A., and Scourfield, J. (2017). Multi-class machine classification of suicide-related communication on twitter. *Online social networks and media*, 2:32–44.
- Burth, D., Godoy, A., Zuben, F. J. V., and Somvir, K. S. (2019). Online social network analysis: A survey of research applications in computer science. *Social network services and Libraries*.
- Castells, M. and Blackwell, C. (1998). The information age: economy, society and culture. volume 1. the rise of the network society. *Environment and Planning B: Planning and Design*, 25:631–636.
- Cohan, A., Desmet, B., Yates, A., Soldaini, L., MacAvaney, S., and Goharian, N. (2018). Smhd: A large-scale resource for exploring online language usage for multiple mental health conditions. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*, pages 1485—1497. Association for Computational Linguistics.
- Fonseca, B. P., Albuquerque, P. C., Zicker, F., and Morel, C. M. (2021). Social network analysis and mining: challenges and applications. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 287–294. SBC.
- Freitas, C. (2013). Sobre a construção de um léxico da afetividade para o processamento computacional do português. *Revista Brasileira de Linguística Aplicada*, 13:1031–1059.

- Hohaus, L. and Berah, E. (1996). Stress, achievement, marriage and social support: effects on the psychological well-being of physicians entering midlife/mid-career. *Psychology and Health*, 11(5):715–731.
- Hott, B., Santos, B. P., Loures, T. C., Benevenuto, F., and Vaz-de Melo, P. O. (2022). Polarização em redes sociais: Conceitos, aplicações e desafios. *Sociedade Brasileira de Computação*.
- Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.
- Lima, A. C. E. and de Castro, L. N. (2019). Tecla: A temperament and psychological type prediction framework from twitter data. *Plos one*, 14(3):e0212844.
- Maia, M., Oliveira, E., and Gallegos, L. (2021). Covid-19 e tweets no brasil: coleta, tratamento e análise de textos com evidências de estados afetivos alterados em momentos impactantes. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 79–90. SBC.
- Martins, L. M., De Alcantara, C., Nardini Barioni, M. C., De Oliveira Júnior, L. C., and Faria, E. R. (2022). A method for analysis of human temperament in contrast to social network data. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, pages 19–27.
- Mikolajczyk, R. T. and Kretzschmar, M. (2008). Collecting social contact data in the context of disease transmission: prospective and retrospective study designs. *Social Networks*, 30(2):127–135.
- Paes, V. J., Araújo, D., Brito, K., and Andrade, E. (2022). Análise de sentimento em tweets relacionados ao desmatamento da floresta amazônica. In *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*, pages 61–72. SBC.
- Plank, B. and Hovy, D. (2015). Personality traits on twitter—or—how to get 1,500 personality tests in a week. In *Proceedings of the 6th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 92–98.
- Ribeiro, J. L. P. (1999). Escala de satisfação com o suporte social (esss).
- Santos, P. D. and Goya, D. H. (2022). Detecção de posicionamento e rotulação automática de usuários do twitter: estudo sobre o embate científico-político no contexto da covid-19. In *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*, pages 49–60. SBC.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12):2544–2558.