

Analysis of Temperament and Perception of Social Support for Online Social Networks

Análise de Temperamento e Percepção de Suporte Social para Redes Sociais *Online*

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. This study proposes a method to contrast the behavior of a user on the online social networks X and Instagram with their perception of social support and temperament, measured through questionnaires. An unsupervised model for classifying the polarity of posts was introduced, achieving superior results compared to the Vader and SentiStrength models. To contrast questionnaire results with user behavior on X and Instagram, analyses of graphs and tables were conducted, along with calculations of Pearson and Point-Biserial correlations, and an ANOVA. The results of the contrast analysis provided valuable insights into understanding the relationship between the real and online worlds, complementing the information obtained through questionnaires.

Keywords. Sentiment Analysis; Text Mining; Online Social Networks; Unsupervised Models; Perception of Social Support; ESSS; Temperament

Resumo. Este estudo propõe um método para contrastar o comportamento de um usuário das redes sociais online X e Instagram com a sua percepção de suporte social e temperamento, medidos por meio de questionários. Foi proposto um modelo não-supervisionado para classificação da polaridade das postagens que obteve resultados superiores aos modelos Vader e SentiStrength. Para contrastar os resultados dos questionários com o comportamento do usuário no X e Instagram, foram realizadas análises de gráficos e tabelas, calculadas medidas de correlação de Pearson e Ponto-Bisserial e conduzida uma ANOVA. Os resultados da análise de contrastes forneceram informações úteis para entender a relação entre o real e o online, complementando as informações obtidas por meio de questionários.

Agradecemos à FAPEMIG (Fundação de Amparo à Pesquisa do Estado Minas Gerais) pelo apoio financeiro a este projeto (Projeto APQ-00226-21).

Palavras-Chave. Análise de Sentimentos; Mineração de Texto; Redes Sociais Online; Modelos Não-Supervisionados; Percepção de Suporte Social; ESSS; Temperamento

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 mediante 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 formação de 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 desordens psicológicas. 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. O estudo do temperamento também desempenha um papel importante na psiquiatria, ao permitir a identificação precoce de riscos e potenciais transtornos psiquiátricos, bem como o desenvolvimento de estratégias de tratamento personalizadas [Rothbart and Derryberry 1981]. Ainda que a percepção de suporte social e o temperamento de uma pessoa sejam temas de estudo na psiquiatria, suas relações com o comportamento de uma pessoa na rede social *online*, ainda não foram exploradas na literatura correlata sobre análise de dados de redes sociais *online*.

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 da literatura propuseram a 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. Em consequência disso, acredita-se que os estudos possam ficar limitados, por poderem 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, levando 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 das rede sociais *online* *X* e *Instagram* com a percepção de suporte social, medida pelo questionário ESSS, e temperamento, medido pelo questionário TEMPS-RIO. 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 posta-

gens, este trabalho propõe uma nova abordagem não-supervisionada baseada em léxicos, evitando assim a criação de bases de dados rotuladas.

Uma versão preliminar do trabalho descrito aqui foi apresentada no artigo [de Alcantara et al. 2023]. O presente artigo estende o trabalho inicial abordando a adição do teste TEMPS-RIO para medir temperamento, assim como dados da rede social online *Instagram*.

As principais contribuições deste artigo podem ser resumidas da seguinte maneira:

- Desenvolvimento de um método original de classificação de polaridade não supervisionado, baseado em léxicos, integrando abordagens existentes com inovações específicas dos autores;
- Criação de bases com dados das redes sociais *online X e Instagram*;
- Representação de um usuário das redes sociais *online X e Instagram*;
- Identificação de padrões no comportamento nas redes sociais *online X e Instagram* em contraste com resultados dos testes de Percepção de Suporte Social e TEMPS-RIO.

O restante do trabalho se estrutura na seguinte maneira: a seção 2 apresenta trabalhos relacionados ao presente estudo, a seção 3 traz conceitos fundamentais deste trabalho, na seção 4 tem-se o método de trabalho, na seção 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*. Dados oriundos do *X* são amplamente utilizados para realizar estudos de variados temas, sendo a maioria envolvendo a análise dos sentimentos expressos 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]. Os trabalhos discutidos nesta seção apresentam algumas semelhanças na abordagem proposta por este trabalho no que tange contraste entre o *online* e fora dele.

O trabalho de [Lima and de Castro 2019] propôs o desenvolvimento de um *framework* para aplicar técnicas de processamento de linguagem natural e classificar a personalidade ou os traços psicológicos de usuários de mídias sociais *online*. Foi utilizada uma base de dados pública do *Twitter* contendo 1,2mi postagens de 1.500 usuários. A rotulação foi realizada através do conteúdo das postagens, onde o próprio usuário se declarou de uma determinada personalidade após realizar o teste de Myers-Briggs [Myers 1962]. A partir dos dados do *Twitter*, os autores criaram uma representação do usuário contendo variáveis comportamentais e gramaticais. Na etapa de modelagem foram testados diversos modelos e combinações de variáveis, cujo objetivo foi, para cada combinação, predizer a personalidade do usuário, dado a sua representação. Ao final, de forma geral, o modelo *Random Forest* apresentou os melhores resultados, atingindo 80% de acurácia aproximadamente. Os resultados foram comparados com outros oito modelos da literatura, obtendo também resultados superiores.

Na pesquisa de [Vioules et al. 2018] foi proposta a utilização de um indicador chamado de IMP para construir um léxico que discriminasse entre quatro classes referentes ao nível de angústia dos usuários do *Twitter* já diagnosticados por profissionais da psicologia. Posteriormente, o IMP foi utilizado com a presença de palavrões, pronomes pessoais e adjetivos de intensidade de cada postagem para encontrar um modelo que predisse a qual classe pertence. Nessa etapa os autores utilizaram os algoritmos *Random Forest*, *SMO*, *Logistic Regression* e *J48* em duas etapas, sendo que na primeira as classes foram agrupadas em dois níveis de intensidade e na segunda foram consideradas as classes do grupo identificado na etapa anterior. Para avaliar a metodologia proposta, os autores utilizaram uma base coletada da rede social *online Twitter*, contendo 500 postagens de uma base maior com 5.446 postagens. Os melhores resultados foram com a combinação dos modelos *J48* na primeira etapa e *SMO* na segunda, atingindo 70% de revocação em média.

No trabalho de [Plank and Hovy 2015] foi realizado um estudo para classificar os usuários da rede social *online Twitter* conforme o MTBI. Para tal, os autores coletaram os dados via API utilizando palavras-chave que identificam cada uma das personalidades, totalizando 1,2 mi postagens de 1.500 usuários. Para criar uma representação dos usuários, os autores trabalharam com dois grupos de variáveis. O primeiro grupo é referente aos textos das postagens, onde utilizaram n-gramas e *tokens* para substituir *links*, *hashtags* e nomes de usuários. O segundo grupo contém variáveis comportamentais dos usuários, como gênero, n.º de postagens, n.º de seguidores, n.º de repostagens, n.º de postagens favoritas e n.º de listas nas quais o usuário aparece. Finalmente, para conseguirem prever a personalidade de cada usuário, foi utilizado o algoritmo Regressão Logística, onde obtiveram uma média de aproximadamente 70% de acurácia.

Apesar de todas as contribuições citadas, constatou-se carência de pesquisas em português que contribuam para a área psiquiátrica ao contrastar o comportamento *online* com resultado de questionários psiquiátricos. Especificamente, até o momento desta pesquisa, não foram encontrados trabalhos que abordem percepção de suporte social nesse contexto. Sendo assim, o presente trabalho visa contribuir para a área da saúde, ao contrastar os questionários estudados neste trabalho com o comportamento nas redes sociais *online* na busca de padrões que auxiliem, por exemplo, em diagnósticos dos pacientes. A Tabela 1 apresenta o resumo dos trabalhos descritos nesta seção.

3. Conceitos Fundamentais

Nesta seção são apresentados os conceitos principais do trabalho. Na Subseção 3.1 é discutido Percepção de Suporte Social. Na Subseção 3.2 é discutido sobre Temperamento.

3.1. Percepção de Suporte Social

A percepção de suporte social é um constructo central na psiquiatria e desempenha um papel crítico na saúde mental e no bem-estar psicológico [Thoits 2011]. Este conceito refere-se à avaliação subjetiva que um indivíduo faz sobre o apoio emocional, instrumental e social disponível em seu ambiente social.

Para mapear a Percepção de Suporte Social, neste trabalho será utilizado o Questionário de Escala de Satisfação com o Suporte Social¹ (ESSS) [Ribeiro 1999], que con-

¹O questionário pode ser visualizado no seguinte link: <https://forms.gle/WvvgyzCAv5xZpaDs9>

Tabela 1. Resumo dos Trabalhos Correlatos — Contexto de Saúde Mental e Mídias Sociais Online

Referência	Questionário	Classificador	Representação do Usuário	Uso de Léxico	Realizada Coleta de Dados
[Lima and de Castro 2019]	Myers-Briggs	Random Forest	Variáveis comportamentais: n.º de postagens, n.º de seguidores, n.º de vezes que foi listado, n.º de seguidos, n.º de vezes que foi favoritado, gênero. Variáveis gramaticais: extraídos dos léxicos LIWC, MRC, sTagger e oNLP	Não	Não
[Vioules et al. 2018]	-	J48 e SMO	O usuário foi representado através de suas postagens, considerando presença de palavrões, pronomes pessoais e adjetivos de intensidade de cada postagem.	Sim	Sim
[Plank and Hovy 2015]	Myers-Briggs	Regressão Logística	Variáveis comportamentais: gênero, n.º de postagens, n.º de seguidores, n.º de repostagens, n.º de postagens favoritas e n.º de listas nas quais o usuário aparece. Variáveis gramaticais: n-gram e uso de tokens para substituir links, hashtags e citação de usuários.	Não	Sim

siste, 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, calculadas segundo as Equações (1), (2), (3) e (4), e o resultado (ESSS) do questionário corresponde a soma de todas as respostas, conforme a Equação (5). Segue uma descrição de cada uma das quatro dimensões e suas equações correspondentes.

- **Satisfação com Amigos:** Avalia o nível de satisfação do indivíduo em relação ao suporte social que recebe de seus amigos. Inclui questões que exploram como o indivíduo se sente em relação à disponibilidade de amigos para conversar, compartilhar experiências e oferecer apoio emocional.
- **Satisfação com Família:** Avalia a satisfação do indivíduo em relação ao suporte social proveniente de membros da família. Inclui questões que investigam como o indivíduo se sente em relação ao apoio prático e emocional de sua família, como auxílio nas atividades diárias, compreensão e apoio emocional.
- **Intimidade:** Avalia a qualidade dos relacionamentos sociais do indivíduo. Inclui perguntas que exploram o grau de intimidade e proximidade que o indivíduo sente em seus relacionamentos sociais.
- **Atividades Sociais:** Avalia o grau de satisfação do indivíduo com suas interações sociais e envolvimento em atividades sociais, como clubes, grupos religiosos ou redes sociais online.

$$satisfacao_amigos = Q1 + Q2 + Q3 + Q4 + Q5 \quad (1)$$

$$intimidade = Q6 + Q7 + Q8 + Q9 \quad (2)$$

$$satisfacao_familia = Q10 + Q11 + Q12 \quad (3)$$

$$atividades_sociais = Q13 + Q14 + Q15 \quad (4)$$

$$ESSS = satisfacao_amigos + intimidade + satisfacao_familia + atividades_sociais \quad (5)$$

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].

3.2. Temperamento

Temperamento pode ser definido como características biologicamente determinadas, sendo estável e relativamente imutável ao longo da vida, que determina o nível básico de reatividade, energia e humor de cada pessoa [Dembińska-Krajewska and Rybakowski 2014]. Cada temperamento possui aspectos positivos e negativos e, além disso, as características do temperamento podem ser construtivas ou destrutivas na qualidade de vida das pessoas.

Para mapear o temperamento, neste trabalho será utilizado o questionário TEMPS-RIO², o qual é uma versão resumida e traduzida para o português do TEMPS-A desenvolvida por [Woodruff et al. 2011], possuindo 45 perguntas de verdadeiro ou falso. O TEMPS-RIO mapeia as cinco personalidades: depressivo, hipertímido, ciclotímido, irritável e ansioso, com a adição do temperamento preocupado, sendo a combinação entre os temperamentos ansioso e depressivo [Karam et al. 2005].

O TEMPS-RIO determina o temperamento do usuário calculando um valor para cada um dos seis temperamentos possíveis. Para tal, são somados os resultados das questões referentes aos temperamentos depressivo (D), ciclotímido (C), irritável (I), hipertímido (H), ansioso (A) e preocupado (P), considerando 1 para as respostas marcadas como "verdadeiro" e 0 para "falso", e multiplicados por um peso, conforme as Equações (6), (7), (8), (9), (10) e (11). O temperamento resultante é o maior valor obtido dentre os temperamentos (Equação (12)), podendo ser mais do que um para cada indivíduo.

Seja T_n a questão de número n do questionário, as equações a seguir calculam o valor referente a cada temperamento:

$$D = (T1 + T2 + T3 + T4 + T5 + T6 + T7 + T8) * 12,5 \quad (6)$$

$$C = (T9 + T10 + T11 + T12 + T13 + T14 + T15 + T16) * 12,5 \quad (7)$$

$$I = (T17 + T18 + T19 + T20 + T21 + T22 + T23 + T24) * 12,5 \quad (8)$$

²O questionário pode ser visualizado no seguinte link: <https://forms.gle/WvvgyzCAv5xZpaDs9>

$$H = (T25 + T26 + T27 + T28 + T29 + T30 + T31 + T32) * 12,5 \quad (9)$$

$$A = (T33 + T34 + T35 + T36 + T37 + T38 + T39 + T40) * 12,5 \quad (10)$$

$$P = (T41 + T42 + T43 + T44 + T45) * 20 \quad (11)$$

O temperamento final é calculado seguindo a Equação (12) a seguir.

$$\text{Temperamento} = \text{MAIOR}(D, C, I, H, A, P) \quad (12)$$

4. Método

Para atingir o objetivo de relacionar percepção de suporte social, medido pelo ESSS, temperamento, medido pelo TEMPS-RIO, e o comportamento de uma pessoa nas redes sociais *online X* e *Instagram*, este trabalho propõe uma estratégia que será descrita nesta seção.

De forma geral, a estratégia é dividida em dois Processos de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases - KDD*), conforme mostrado nas Figuras 1 e 2.

No primeiro, chamado de "classificação das postagens", o objetivo é criar e validar um método para classificação de polaridade de cada postagem de usuários de redes sociais *online* que será utilizado posteriormente para construir a representação de um usuário para cada rede social *online*. Já no segundo, chamado de "contraste entre o real e o *online*", a finalidade é, a partir de representações criadas para os usuários de cada rede social *online*, contrastar o comportamento do real, medido por meio dos questionários de suporte social e de temperamento, com o *online*.

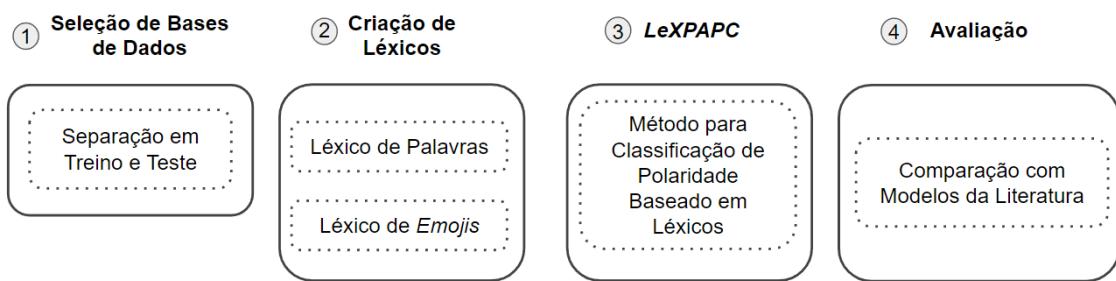


Figura 1. Representação Gráfica do Processo de Classificação das Postagens do Método.

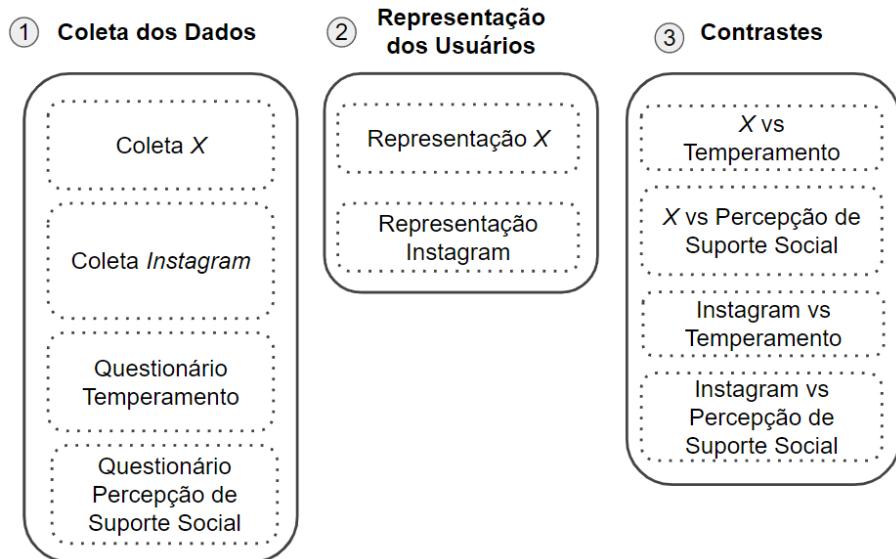


Figura 2. Representação Gráfica do Processo de Contraste entre o Real e o *Online* do Método.

4.1. Classificação das Postagens

O objetivo do primeiro processo de KDD, mostrado na Figura 1, chamado de "classificação das postagens" é criar e avaliar um método que consiga classificar, de forma não supervisionada, a polaridade (1, 0 ou -1) das postagens dos usuários das redes sociais *online* *X* e *Instagram*.

4.1.1. Seleção de Bases de Dados

Para a realização desta etapa do trabalho foram escolhidas bases de dados já previamente utilizadas na literatura em trabalhos anteriores que obedecessem os seguintes critérios: fossem rotuladas com as classes positivo (1), negativo (-1) ou neutro (0), tivessem sido extraídas de alguma rede social *online* e estivessem em português.

Após coletadas as bases, cada uma foi separada em duas seguindo a proporção de 70% para a criação do método para classificação de polaridade baseado e léxico e 30% para a avaliação do método.

4.1.2. Criação dos Léxicos

Para ser possível a criação de um método não supervisionado para classificação de polaridade baseado em léxico é necessária a utilização de léxicos, os quais são tabelas contendo pelo menos duas colunas, uma indicando a sequência de caracteres e a outra o seu peso atribuído.

Neste trabalho foram criados dois léxicos, um de palavras e um de *emojis*. Visando determinar uma estratégia para a criação dos léxicos, primeiramente, fez-

se uma análise dos algoritmos Vader [Hutto and Gilbert 2014] e SentiStrength (SS) [Thelwall et al. 2010], onde é possível notar que a abordagem de atribuir um peso que reflete a intensidade de cada uma das palavras contidas em uma postagem podem acarretar alguns problemas.

Para exemplificar, 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 tinha 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 o módulo do peso da palavra "caos" é maior do que da palavra "alegria".

Outra questão que pode ocorrer é na atribuição dos pesos nos léxicos. Uma abordagem comum é a votação, na qual tem-se um grupo de pessoas que determina o peso de cada termo e constrói-se o léxico a partir disso.

Nesse caso, apesar de ser uma estratégia que atinge bons resultados, acredita-se que podem acontecer dois problemas, o primeiro é que os pesos dos termos podem ter o viés do contexto da base de dados analisada, e o segundo é referente ao viés de quem votou nos pesos.

Objetivando contornar os problemas apontados, foi adotada a estratégia descrita a seguir:

- 70% das bases foram utilizadas na criação do método para classificação de polaridade, assim como dos léxicos utilizados no método.
- A estratégia é igual para o léxico de palavras e para o léxico de *emojis*.
- Foram considerados termos de até três palavras (ex.: chutar cachorro morto). *Emojis* não foram agrupados, sendo um peso para cada *emoji*.
- Cada termo e *emoji* pode receber os pesos 1 ou -1 . Por conta disso, registros classificados como "neutro"(0) do banco de dados foram desconsiderados.
- Os léxicos começam vazios e vão sendo preenchidos durante um processo iterativo que será descrito nas próximas etapas.
- Em cada iteração, a base de teste é classificada pelo método de classificação de polaridade e as classificações incorretas são separadas.
- Das classificações incorretas, são separados os principais termos positivos e os negativos. Que são aqueles termos que aparecem mais em cada rótulo, isto é, os principais termos positivos são aqueles que aparecem mais nos registros positivos classificados incorretamente (Falso Negativo). A mesma lógica vale para os principais termos negativos. Esses termos não são conhecidos pelo algoritmo até então.
- O primeiro termo da lista dos principais positivos e negativos são candidatos a serem adicionados ao léxico.
- O léxico anterior é substituído por um novo léxico, sendo o léxico da iteração anterior com os termos adicionais. O novo léxico é submetido ao método de

classificação de polaridade e as medidas de precisão, acurácia e F1 são calculadas. Caso haja, de forma geral, melhoria nos resultados, o novo léxico com a adição dos termos testados é mantido, caso contrário eles são desconsiderados na próxima interação.

4.1.3. Método para Classificação de Polaridade Baseado em Léxico - LeXPAPC

O método não supervisionado para classificação de polaridade baseado em léxico e regras de linguagem, desenvolvido especificamente neste trabalho, foi denominado ***Lexical Portuguese Approach for Polarity Classification*** (LeXPAPC). Este método foi concebido integrando abordagens e técnicas existentes com inovações próprias dos autores, resultando em um processo único e adaptado ao contexto da língua portuguesa.

A partir da entrada de uma base de dados contendo textos, o LeXPAPC executa todas as etapas de processamento dos dados, incluindo pré-processamento, representação e classificação da polaridade de cada registro. O fluxo de execução do LeXPAPC é mostrado na Figura 3, detalhando as etapas criadas pelos autores e as adaptações feitas a partir de técnicas consagradas na literatura. Cada passo do método está descrito a seguir.

1. **Entrada de dados:** O método foi desenvolvido para aceitar qualquer base de dados textual, principalmente com características de redes sociais *online*, no qual cada linha, que será chamado de registro, contém, além do texto, *emojis*, *links*, citações de usuários, compartilhamento de arquivos, etc.;
2. **Tratamento dos dados:** Esta corresponde às etapas de pré-processamento e representação dos dados. Para o pré-processamento as seguintes etapas são executadas: todos os caracteres são colocados em caixa-baixa, são removidos acentos, pontuação, links, citação de usuários, imagens, *Gifs*, vídeos e palavras de parada (*stopwords*). A representação dos textos é através do N-grama, considerando $n = 1, 2$ e 3 ;
3. **Separação de Vetores de Entidades:** Cada registro é separado em três vetores chamados de entidades, são eles: **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 vazio. Caso os três sejam vazios, esse registro não passará para as próximas etapas e não será classificado;
4. **Consulta nos Léxicos:** Cada elemento de cada vetor de entidade é consultado em seu léxico correspondente, isto é, textos e *hashtags* são consultados no léxico de palavras e *emojis* são consultados no léxico de *emojis*. Caso haja correspondência, o valor indicado nos léxicos é atribuído ao termo, podendo ser 1 ou -1;
5. **Regras Gramaticais:** Assim como no Vader, entende-se que termos que expressam adversão podem ser tradados separadamente [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. A classificação final será

um destes valores resultantes, que será "positivo" caso o valor seja positivo, "negativo" caso o valor seja negativo, "neutro" caso a soma seja zero e pelo menos um dos valores utilizados na soma difere de zero e "não conclusivo" nos outros casos. Para reduzir a incidência de classificações "não conclusivo", os modelos Vader [Hutto and Gilbert 2014] e SentiStrength (SS) [Thelwall et al. 2010] também foram utilizados para classificar os registros. Para determinar a classificação final, as seguintes estratégias são consideradas:

- **Lexicografia:** Como foi mostrado no Item 3, um ou mais vetores de entidade podem não existir para determinado registro. A estratégia aqui é a classificação final receber o valor resultante de um dos vetores de entidades, obedecendo à hierarquia, *emoji*, *hashtag* e, por último, texto.
- **Funil:** Considera-se como classificação final o primeiro modelo que obter classificação que não seja "não conclusivo", obedecendo a seguinte hierarquia: LeXPAPC, Vader e SS. Sendo que, para o LeXPAPC, a classificação é feita utilizando a estratégia de "Lexicografia" explicada acima.
- **Votação:** Considera-se como classificação final o voto da maioria dos três modelos: LeXPAPC, Vader e SS. Sendo que, para o LeXPAPC, a classificação é feita utilizando a estratégia de "Lexicografia" explicada acima.

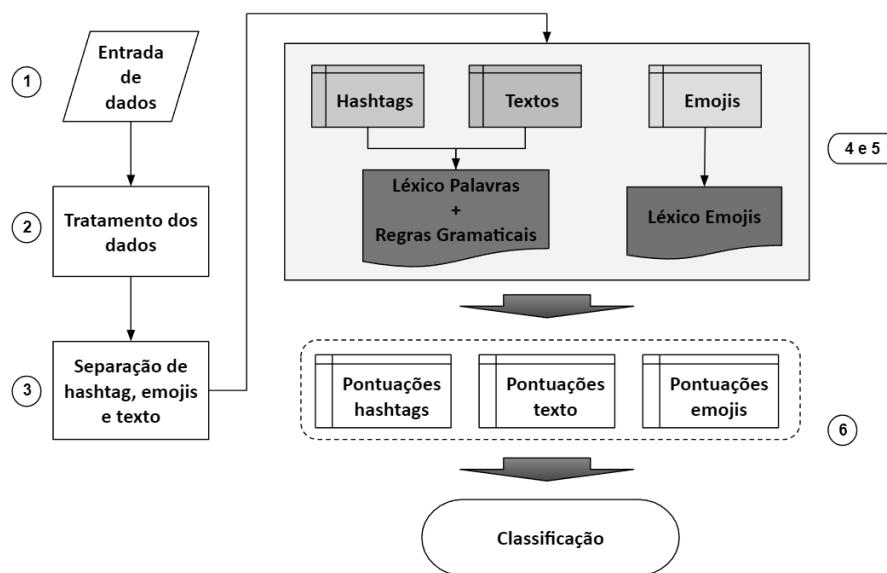


Figura 3. Representação gráfica do método LeXPAPC. Adaptado de [de Alcantara et al. 2023].

Na Figura 4, tem-se um exemplo de como um registro se comporta ao ser submetido ao LeXPAPC. 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, sendo -4 para o texto e 1 para os emojis. A classificação final foi feita utilizando a estratégia de lexicografia, ficando com o valor dos *emojis* e resultando ”positivo”.

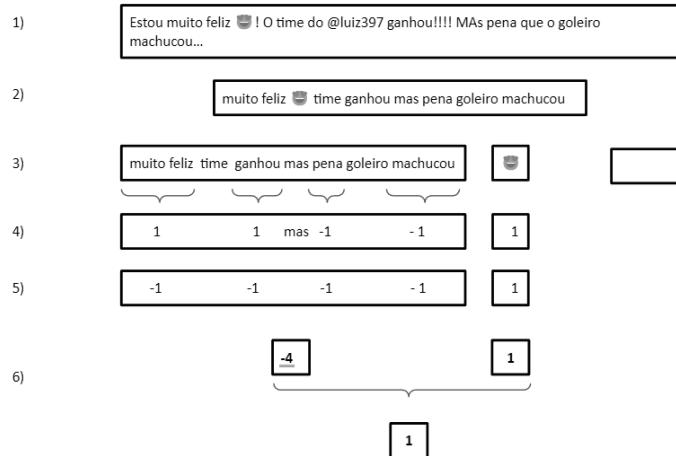


Figura 4. Exemplo do comportamento de um registro submetido ao LeXPAPC.

4.1.4. Avaliação

O LeXPAPC foi avaliado nas bases de testes, que correspondem a 30% do todo. São calculadas as medidas de acurácia, precisão, F1 e abrangência após aplicação dos métodos LeXPAPC, considerando as três estratégias de classificação descritas anteriormente, Vader e SentiStrength. A avaliação se dá comparando os resultados das medidas listadas de cada método para cada uma das bases mostradas nesta Seção.

4.2. Contraste entre os Dados Obtidos a Partir de Questionários e Dados das Redes Sociais *Online*

O segundo processo de KDD, mostrado na Figura 2, objetiva contrastar o comportamento do real, medido através dos questionários de percepção de suporte social e de temperamento, com o *online*, obtido através da representação dos usuários de cada rede social *online* utilizada neste trabalho.

4.2.1. Coleta dos Dados

Para a realização deste trabalho criou-se bases próprias a partir de dados de voluntários que são discentes da Universidade Federal de Uberlândia (UFU). Com este objetivo, um projeto foi submetido e aprovado pelo Comitê de Ética em Pesquisas com Seres Humanos — CAAE 35533420.2.0000.5152 — que autoriza a criação dessas bases.

A coleta possui três etapas: coleta de dados do *Instagram*, coleta de dados do *X* e questionários de temperamento e suporte social. Os dados são coletados apenas da-

queles que aceitaram participar da pesquisa de forma voluntária e assinaram o Termo de Compromisso Livre e Esclarecido. Para ser possível a coleta, foi desenvolvido um *site*³.

O método de amostragem utilizado foi o não probabilístico de amostragem por conveniência, no qual são selecionados para participar da amostra aqueles que têm disponibilidade de participar da pesquisa, considerando a facilidade de acesso.

A coleta das redes sociais *online* resultou em duas bases, conforme listado a seguir:

- **BASE-X:** Base resultante da coleta de dados da mídia social *online X*, possuindo 15.711 registros de 39 respondentes. As colunas desta base são: quantidade de seguidos, quantidade de seguidores, quantidade de curtidas da postagem, quantidade de postagens, texto da postagem, data e hora da postagem, uso de *media* na postagem.
- **BASE-INSTA:** Base resultante da coleta de dados da mídia social *online Instagram*, possuindo 1.860 registros de 85 respondentes. As colunas desta base são: quantidade de seguidos, quantidade de seguidores, quantidade de curtidas da postagem, quantidade de postagens, texto da postagem, data e hora da postagem.

O total de respondentes considerando as duas bases é de 108, pois existem 16 pessoas com contas no *X* e no *Instagram* ao mesmo tempo, as demais possuem conta apenas em uma das duas redes.

Somado às duas bases listadas acima, também tem-se como base de dados as respostas dos questionários TEMPS-RIO e de Percepção de Suporte Social. A seguir é explicado como os dados coletados foram utilizados para a criação da representação dos usuários.

4.2.2. Representação dos Usuários

A partir dos dados coletados pelo processo de coleta, este trabalho propõe duas representações de usuário, sendo uma para o *X* e outra para o *Instagram*, por meio de conjuntos de atributos, separados entre **midiáticas**, que são referentes a pessoa que é dona do perfil e **questionários**, que são resultantes dos questionários de temperamento e de percepção de suporte social. A seguir, são descritos cada conjunto de atributos que compõem as representações dos usuários.

Variáveis Midiáticas:

As variáveis midiáticas são obtidas a partir do perfil de cada usuário e em cada mídia social *online*. Portanto, elas refletem o padrão de comportamento de cada usuário em uma mídia social *online*. A seguir tem-se a lista das variáveis midiáticas e como elas foram criadas para cada mídia social *online*.

As variáveis midiáticas são:

³<https://www.socialtemp.facom.ufu.br/>

- **media-curtidas:** Média de curtidas por postagem de cada usuário. Calculada dividindo a quantidade total de curtidas entre todas as postagens pela quantidade de postagens de cada usuário;
- **quantidade-postagens:** Quantidade de postagens que o usuário possui em seu perfil. Calculada contando a quantidade de postagens de cada usuário;
- **perc-postagens-positivas:** Após a aplicação do método LeXPAPC, foi calculada a porcentagem de postagens positivas para cada usuário;
- **perc-uso-media:** Variável binária somente para o *X* que indica se há presença de imagens ou vídeos nas postagens. Posteriormente foi calculada a porcentagem da mesma.
- **perc-uso-emoji:** Variável somente para o *Instagram* que representa a porcentagem de postagens contendo *emojis* no texto. Calculada após a criação de uma variável binária indicando a existência de *emoji* em cada texto. A soma dessa variável binária indica a quantidade de postagens contendo *emojis*, dividida pela quantidade total de postagens;
- **perc-final-semana:** Variável somente para o *Instagram* que representa a porcentagem de postagens feitas aos finais de semana. Calculada após a criação de uma variável binária que identifica se determinada postagem foi feita sábado ou domingo. A soma dessa variável binária indica a quantidade de postagens aos finais de semana, dividida pela quantidade total de postagens;
- **proporcao-seguidos-seguidores:** A relação entre a quantidade de seguidos e seguidores. Calculada dividindo a quantidade de seguidos (*following*) pela quantidade de seguidores (*followers*);

Variáveis Resultantes dos Questionários:

As variáveis resultantes dos questionários são calculadas aplicando os cálculos mostrados na Seção 3. Não há diferença dessas variáveis nas diferentes redes sociais *online*. A seguir tem-se a lista das variáveis referentes ao questionário de temperamento e de percepção de suporte social calculadas.

- **genero:** Autodeclarado pelos usuários nas respostas do questionário aplicado, sendo 1 para masculino e 0 para feminino;
- **satisfacao-amigos:** Calculado conforme descrito na Equação (1);
- **intimidade:** Calculado conforme descrito na Equação (2);
- **satisfacao-familia:** Calculado conforme descrito na Equação (3);
- **atividades-sociais:** Calculado conforme descrito na Equação (4);
- **temperamento:** Calculado conforme descrito na Equação (12). Para este trabalho, em caso de empate foi considerado o primeiro temperamento que apareceu nos cálculos intermediários, mostrados nas Equações (6), (7), (8), (9), (10) e (11).

4.2.3. Representação do Usuário do *X* e do *Instagram*

Utilizando as variáveis mostradas em 4.2.2 constrói-se, para cada rede social *online* considerada neste trabalho, as representações dos usuários conforme mostrado nas Figuras 5 e 6. Para facilitar a descrição dos resultados experimentais, foram criados apelidos para cada variável que, a partir daqui, serão utilizados para referenciá-las.

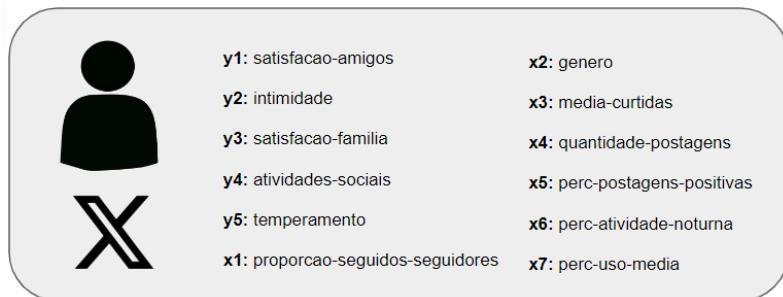


Figura 5. Representação de um Usuário do X.

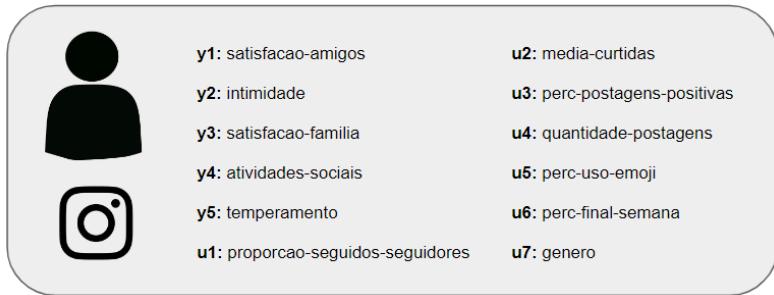


Figura 6. Representação de um Usuário do Instagram.

4.3. Contrastes

Para contrastar o comportamento online com o real são utilizados testes estatísticos, medidas de correlação e análise de gráficos e tabelas. Os testes estatísticos escolhidos foram Análise de Variância (ANOVA) com o teste de Tukey, ao nível de 5% de significância. As medidas de correlação utilizadas foram as de Pearson e Ponto-Bisserial. Os gráficos utilizados foram *box plot*, histograma e gráfico de pizza.

Em relação aos testes, foram consideradas apenas as dimensões que compõem o ESSS (atividades sociais, intimidade, satisfação família e satisfação amigos) nas análises, e não o seu valor final. Em alguns casos foi necessária a criação de variáveis binárias relacionadas aos temperamentos, indicando se determinado indivíduo é (1) ou não (0) de determinado grupo de temperamento. Seguindo a seguinte lista: z1 - ciclotímido, z2 - hipertímido, z3 - preocupado, z4 - irritável, z5 - ansioso e z6 - depressivo

A seguir tem-se uma lista das questões de pesquisa abordadas, assim como a combinação das técnicas citadas acima utilizadas para respondê-las:

1. **Pergunta:** Qual o comportamento geral dos respondentes em relação aos testes?
Estratégia: Análise de gráficos de pizza, *box plot* e histograma.
2. **Pergunta:** Existe diferença nas dimensões da percepção de suporte social e os diferentes temperamentos? **Estratégia:** ANOVA para avaliar se há alguma diferença com teste de Tukey para identificar onde estão as diferenças, caso haja.
3. **Pergunta:** Em cada rede social *online* considerada neste trabalho, existe relação no comportamento de cada variável midiática criada para representar um usuário e cada dimensão do teste de percepção de suporte social? **Estratégia:** Cálculo

dos coeficientes de correlação de Pearson, para pares de variáveis contínuas, e Ponto-Bisserial, para uma variável discreta e uma contínua, entre as variáveis de representação e as dimensões do ESSS. Posteriormente, construção de tabelas com os coeficientes e análise do sinal (direção) e força das correlações.

4. **Pergunta:** Em cada rede social *online* investigada neste trabalho, existe relação no comportamento das variáveis midiáticas criadas para representar um usuário e os resultados do teste TEMPS-RIO? **Estratégia:** Cálculo dos coeficientes de correlação de Ponto-Bisserial entre as variáveis de representação e as variáveis binárias que representam cada temperamento. Posteriormente, construção de tabelas com os coeficientes e análise do sinal (direção) e força das correlações.

5. Experimento e Resultados

Esta seção descreve os experimentos executados neste trabalho assim como seus resultados. O conteúdo está separado em duas etapas, sendo a primeira referente à avaliação do LeXPAPC e a segunda apresenta o contraste dos resultados dos questionários de Percepção de Suporte Social e TEMPS-RIO com as redes sociais *online* X e Instagram.

5.1. Avaliação do LeXPAPC

Para a criação e avaliação do LeXPAPC foram selecionadas três bases da literatura segundo os critérios mostrados na Subseção 4.1.1. A seguir são apresentadas cada base, 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* X e coletada pela API do site [Silva et al. 2019]. A coleta foi por palavras-chave, extraídas do TeP 2.0 ⁴. A rotulação foi realizada mediante 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 [Brum and Volpe Nunes 2018] é uma base de dados referente ao X 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.
- **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 xxxxxxxx.x.xxxx.xxxx, 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 a partir da comparação dos resultados obtidos após sua aplicação nas bases Tash-pt, TweetSentBR e TEMPS-RIO em contraste com os modelos da literatura Vader e SS. Para as previsões do LeXPAPC foram consideradas as estratégias

⁴<http://www.nilc.icmc.usp.br/tep2/index.htm>

de Lexicografia, Votação e Funil mostradas em 4.1.3. Os resultados estão dispostos na Tabela 2 e na Figura 7.

Primeiramente, na Figura 7 tem-se as medidas de abrangência em barras e F1 na linha pontilhada resultantes da aplicação do LeXPAPC com a estratégia de lexicografia para classificação dos registros nas bases Tash-pt, TweetSentBR e TEMPS-RIO. Observando a figura, nota-se que em todas as bases partes das suas instâncias receberam uma classificação negativa ou positiva. Isso significa que menos de 50% dos registros receberam uma classificação positiva, negativa ou neutra.

Esse comportamento é esperado quando se utiliza apenas o LeXPAPC para classificar uma base de dados, isto é, quando não se utiliza as estratégias de Funil ou Votação mostradas na Seção 4.1.3. O LeXPAPC só classifica um registro quando encontra algum termo ou *emoji* em um dos léxicos criados, sendo assim, a tarefa de classificar uma base por completo é difícil para o algoritmo criado, uma vez que precisaria listar uma quantidade enorme de termos e *emojis*, o que acredita-se que acarretaria perda de qualidade do classificador.

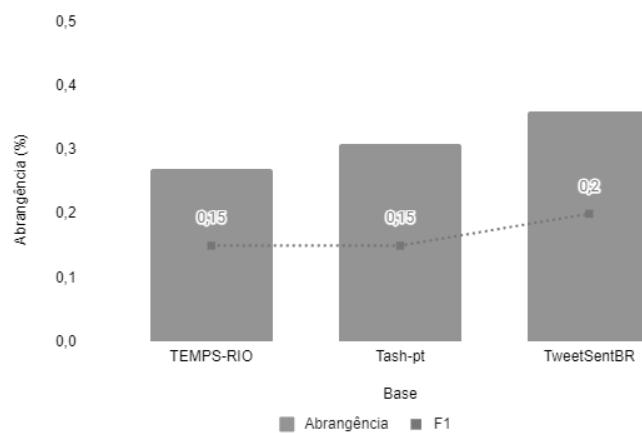


Figura 7. Abrangência e F1 do LeXPAPC com a Estratégia de Classificação Lexicografia para as Bases Tash-pt, TweetSentBR e TEMPS-RIO.

Na Tabela 2 têm-se as métricas de Acurácia, Precisão, F1 e Abg para todas as bases e métodos considerados, destacando, para cada base, os maiores valores de cada medida, exceto Abrangência. Observando a tabela, tem-se que as estratégias de Votação e Funil apresentaram bons resultados nas medidas de avaliação, ficando acima em todas na Base TEMPS-RIO, na Base TweetSentBR apenas a Acurácia não apresentou o maior valor e na Base Tash-pt ficou empatado com os melhores resultados na segunda casa decimal com a estratégia de Votação, mas com a abrangência em 100%.

Com base nos resultados mostrados, pode-se concluir que o método LeXPAPC com a estratégia Funil apresentou os melhores resultados nas medidas de avaliação de forma geral, além de atingir 100% de abrangência, mostrando ser um bom classificador não supervisionado para polaridade. Portanto, o LeXPAPC com a estratégia Funil foi adotado como classificador utilizado para classificar as postagens a fim de criar a variável perc-postagens-positivas, mostrada em 4.2.2.

Tabela 2. Medidas de Acurácia, Precisão, F1 e Abrangência dos Métodos Vader, SS e LeXPAPC (Hierarquia, Votação e Funil) para as Bases Tash-pt, TweetSentBR e TEMPS-RIO.

Base	Método	Acurácia (%)	Precisão (%)	F1 (%)	Abg (%)
Tash-pt	VADER	0,4529	0,4615	0,4646	1
	SS	0,4385	0,4323	0,4316	1
	Lexicografia	0,1673	0,1343	0,1456	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
	Lexicografia	0,2797	0,2011	0,1989	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
	Lexicografia	0,1444	0,1155	0,1461	0,27
	Votação	0,5556	0,5173	0,5090	1
	Funil	0,5697	0,5360	0,5528	1

5.2. Contraste entre o Real e o *Online*

O contraste do comportamento entre real e o *online*, conforme descrito em 4.3, foi direcionado por uma série de perguntas de pesquisa. Pela característica da amostragem no processo de coleta de dados mostrado na Seção 4.2, assim como a quantidade baixa de respondentes com perfil em alguma rede social *online*, não serão realizadas inferências estatísticas, mas sim análises dos respondentes da pesquisa, como sendo o universo de estudo. A seguir é apresentada a discussão acerca de cada uma das perguntas de pesquisa.

5.2.1. Qual o Comportamento Geral dos Respondentes em Relação aos Testes?

Para entender o comportamento dos respondentes em relação aos seis temperamentos construiu-se um histograma, como pode ser visto na Figura 8. Ao analisar a figura, percebe-se um desequilíbrio na distribuição dos temperamentos, onde os temperamentos irritável, depressivo e ansioso aparecem pouco, enquanto as outras personalidades apresentam números maiores de respondentes.

Para avaliar os resultados do teste de percepção de suporte social, escolheu-se observar as dimensões intimidade, satisfação amigos, satisfação família e atividades sociais, que fazem parte do cálculo do ESSS.

Na Figura 9 têm-se os gráficos box-plot de cada dimensão do ESSS. Observando-a, é possível notar que satisfação_amigos apresenta maiores valores, e atividades_sociais os menores. Também é importante salientar que os valores no geral foram baixos considerando o máximo de cada dimensão, o que pode ser um reflexo do período no qual foi realizada a pesquisa, que coincidiu com o período de pandemia do COVID-19.

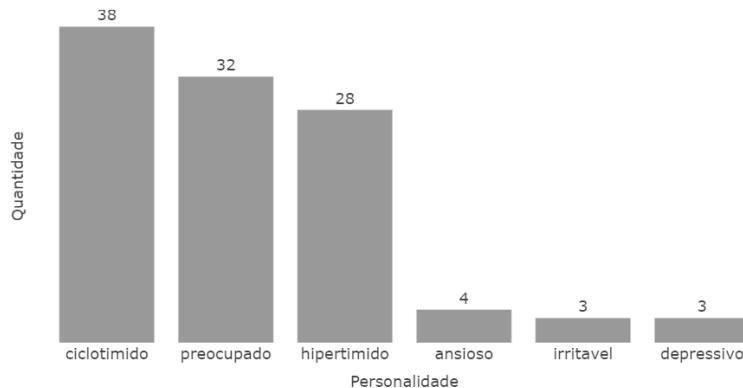


Figura 8. Distribuição dos Respondentes em cada Temperamento.

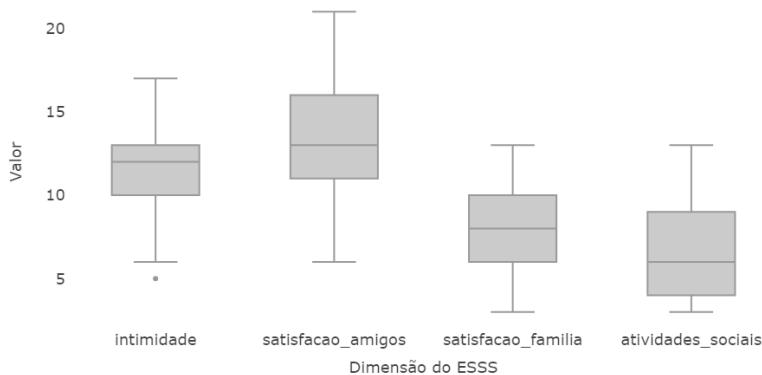


Figura 9. Comparação de Distribuições por Dimensão do ESSS.

5.2.2. Existe Diferença nas Dimensões do ESSS e os Diferentes Temperamentos?

Para comparar os resultados dos dois questionários foram realizadas análises de variância (ANOVA), ao nível de 5% de significância, considerando cada dimensão do ESSS em contraste com os seis temperamentos do TEMPS-RIO.

Os resultados estão dispostos na Tabela 3, onde se tem, para cada dimensão do ESSS, os valores da estatística do teste F e seu p-valor. Na tabela observa-se que os p-valor de atividades sociais e intimidade ficaram abaixo do nível de significância, o que indica que existem evidências estatísticas o suficiente para afirmar que existe alguma diferença dos resultados de ESSS para algum par de temperamentos.

Como existem diferenças significativas pela ANOVA, o próximo passo é localizá-las. Para tal, executou-se o teste de Tukey, também ao nível de 5% de significância, comparando os temperamentos entre si em relação à intimidade e atividades sociais.

Os resultados dos testes foram significativos para os pares (depressivo, hi-

Tabela 3. Resultados da Análise de Variância entre os Temperamentos e as Dimensões do ESSS.

	Estatística do Teste F	p-valor
atividades_sociais	2,7849	0,0212
intimidade	3,6760	0,0042
satisfacao_familia	2,0057	0,0840
satisfacao_amigos	1,4885	0,2001

pertímidos) em relação à intimidade e (hipertímidos, preocupados) em relação à atividades_sociais, com p-valores de 0,0105 e 0,0231, respectivamente. O que indica que, em média, hipertímidos possuem valores maiores de intimidade do que depressivos e que preocupados possuem valores maiores de atividades sociais do que hipertímidos.

5.2.3. Existe Relação no Comportamento de Cada Variável Textual e Comportamental e Cada Dimensão do ESSS? Existe Relação no Comportamento de Cada Variável Textual e Comportamental e os Temperamentos?

Para entender a relação do comportamento de cada variável textual e comportamental com os resultados dos dois testes aplicados utiliza-se as bases criadas a partir da coleta dos dados das redes sociais *online X* e *Instagram*, as bases BASE-X e BASE-INSTA. As análises estão separadas para cada uma das bases.

- **Análises para a BASE-X**

Para avaliar o comportamento de cada variável textual e comportamental (xi) em relação às dimensões do ESSS (yi) e os temperamentos (zi) foram calculados os Coeficientes de Correlação de Pearson e Ponto-Bisserial. O Coeficiente de Correlação de Pearson foi calculado para os pares $(x1, y1)$, $(x1, y2)$, $(x1, y3)$, $(x1, y4)$, $(x3, y1)$, $(x3, y2)$, $(x3, y3)$, $(x3, y4)$, $(x4, y1)$, $(x4, y2)$, $(x4, y3)$, $(x4, y4)$, $(x5, y1)$, $(x5, y2)$, $(x5, y3)$, $(x5, y4)$, $(x6, y1)$, $(x6, y2)$, $(x6, y3)$, $(x6, y4)$, $(x7, y1)$, $(x7, y2)$, $(x7, y3)$ e $(x7, y4)$. Para os demais foi calculado o Coeficiente de Correlação Ponto-Bisserial. A descrição de cada variável xi , yi e zi está na Seção 4.2.

Nas Tabelas 4 e 5 tem-se os valores dos coeficientes de correlação para cada par de variável. As cores das células indicam a direção da correlação, sendo cinza para correlações negativas e branco para positivas.

Tabela 4. Correlações de Pearson e Ponto-Bisserial e o Teste de Percepção de Suporte Social para a BASE-X.

	x1	x2	x3	x4	x5	x6	x7
y1	-0,1153	0,0143	0,0669	0,0969	-0,2313	-0,0585	-0,2692
y2	-0,0851	-0,0214	0,1974	0,1809	0,2341	0,4292	0,3134
y3	0,0328	-0,2998	-0,1413	0,0298	0,0512	0,3956	-0,0618
y4	0,0368	-0,0113	0,2043	-0,1519	-0,1058	0,3241	0,1110

Tabela 5. Correlação Ponto-Bisserial e o Teste TEMPS-RIO para a BASE-X.

	x1	x2	x3	x4	x5	x6	x7
z1	-0,1740	-0,3121	-0,0294	-0,0223	-0,0590	-0,1889	0,0208
z2	-0,1011	0,3967	0,0299	-0,1376	0,2384	0,2204	0,0979
z3	0,3485	0,1060	0,0361	-0,0712	-0,0813	-0,0935	-0,1260
z4	0,1141	-0,2036	-0,0403	0,6950	-0,1088	0,0504	-0,0827
z5	-0,1287	-0,2036	-0,0403	-0,0634	-0,1556	0,1410	0,0456

Observando as Tabelas 4 e 5 é possível obter diversas conclusões. A seguir tem-se uma lista de algumas que foram consideradas pertinentes:

- A relação entre a quantidade de usuários que o respondente segue pela quantidade de usuários que seguem o respondente no X (x2), na Tabela 5, é negativa para os temperamentos ciclotímido (z1), preocupado (z3) e ansioso (z5). Isso quer dizer que para esses três temperamentos, ter uma quantidade menor de seguidores do que usuários que segue, é uma característica marcante.
- A única correlação forte dentre as duas tabelas é, na Tabela 5, entre o temperamento irritável (z4) e a quantidade de postagens (x4) que o respondente possui. Além de ser a única correlação positiva na tabela para essa variável. A mesma variável x4 possui correlação positiva com todas as dimensões do ESSS, exceto atividades sociais (y4), na qual, quanto menor a quantidade de postagens, maiores os valores de atividades sociais.
- Percentual de atividade noturna (x6), na Tabela 4, apresenta no geral as correlações mais fortes, sendo negativo apenas para satisfação amigos (y1), indicando que quem posta mais durante a noite possui menores valores de satisfação amigos. Já na Tabela 5, as correlações não são tão fortes para essa variável, que apresenta valores negativos para os temperamentos ciclotímido (z1) e preocupado (z3), indicando que pessoas desses dois temperamentos postam mais durante o dia.

Além dos pontos listados acima, a partir das Tabelas 4 e 5, pode-se traçar os comportamentos dos respondentes em relação às duas pesquisas ao observar os sinais das correlações. A seguir tem-se dois exemplos, um para cada tabela, de como fazer a interpretação dos sinais e chegar em comportamentos:

- Aqueles possuem altos valores de intimidade (y2) seguem menos usuários do que tem usuários os seguindo (x1), são mulheres (x2) e possuem valores mais baixos de intimidade. Já quem possui muito curtidas nas postagens (x3), possui uma quantidade grande de postagens (x4), majoritariamente positivas (x5), postam mais durante a noite (x6) e costumam usar imagens e vídeos nas postagens (x7).
- Comparando com os outros temperamentos, os preocupados (z3) tem tendência de possuírem mais seguidores (x1), serem do gênero masculino (x2), possuírem mais curtidas nas postagens (x3), terem poucas postagens (x4), o conteúdo das postagens ter sentimento majoritariamente negativo (x5), postarem mais durante o dia (x6) e usarem poucas imagens e vídeos nas postagens (x7).

- **Análises para a BASE-INSTA**

As análises na BASE-INSTA seguem a mesma lógica do que foi mostrado para a BASE-X. O Coeficiente de Correlação Ponto-Bisserial foi calculado para todos os pares de variável, uma vez que as variáveis z_i são dicotômicas. Nas Tabelas 6 e 7 tem-se os valores dos coeficientes de correlação para cada par de variável. As cores das células nas tabelas indicam a direção da correlação, sendo cinza para correlações negativas e branco para positivas.

Tabela 6. Correlações de Pearson e Ponto-Bisserial e o Teste de Percepção de Suporte Social para a BASE-INSTA.

	u1	u2	u3	u4	u5	u6	u7
y1	0,0400	0,0404	-0,1256	-0,1364	-0,2288	-0,1506	0,0750
y2	0,0933	0,0990	-0,0601	0,0232	0,1373	0,1022	0,0021
y3	0,0612	0,0677	0,1094	0,0468	0,0024	0,0745	0,0833
y4	-0,1149	-0,1277	-0,0723	-0,1129	-0,1504	-0,0896	0,0482

Tabela 7. Correlação Ponto-Bisserial e o Teste TEMPS-RIO para a BASE-INSTA.

	u1	u2	u3	u4	u5	u6	u7
z1	0,0374	0,1556	0,0651	0,1450	-0,2040	0,0223	-0,0086
z2	-0,0995	0,0755	-0,0060	-0,0435	0,2063	-0,0648	0,1507
z3	-0,0616	-0,1540	-0,1587	-0,0895	-0,0941	0,0735	-0,0415
z4	0,0354	-0,0699	-0,0169	-0,0173	0,0793	-0,1035	0,1146
z5	-0,0095	0,0059	0,0472	0,0012	0,1022	-0,0353	-0,1727
z6	0,2710	-0,0991	0,1783	-0,0231	0,0770	0,0284	-0,1486

A seguir têm-se algumas conclusões pertinentes em relação às Tabelas 6 e 7:

- Na Tabela 6, o gênero dos respondentes (u7), possui correlação positiva com todas as dimensões do ESSS, o que indica que pessoas do gênero masculino possuem maiores valores de correlação com as dimensões do ESSS do que as mulheres. Além disso, na mesma tabela, atividades sociais (y4) possui correlação negativa com todas as outras variáveis. Este fato corrobora com o mostrado na Figura 9, na qual os valores de atividades sociais são os menores no geral.
- Na Tabela 7 tem-se que o temperamento preocupado (z3) possui correlação negativa com todas as variáveis, exceto u6, que indica que pessoas preocupadas postam mais aos finais de semana.

Assim como foi feito para a BASE-X, a partir das Tabelas 6 e 7, pode-se traçar comportamentos em relação às duas pesquisas observando os sinais das correlações. Na sequência são listados dois exemplos, um para cada tabela, da interpretação dos comportamentos:

- Quem segue menos usuários do que a quantidade de usuários seguindo (u1), tem muito curtidas nas postagens (u2), posta conteúdos mais positivos (u3), faz muitas postagens (u4), faz muito uso de *emoji* (u5), posta mais nos finais de semana (u6), é do gênero masculino (u7) e possui valores mais altos de satisfação família (y3).

- Depressivos (z6) possuem mais seguidores do que seguem outros usuários (u1), não têm muito curtidas nas postagens (u2), postam conteúdos geralmente positivos (u3), não postam muito (u4), usam muito *emojis* (u5), postam mais aos fins de semana (u6) e são principalmente do gênero feminino (u7).

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 de redes sociais *online* (LeXPAPC) e a realização de uma análise do contraste entre o comportamento nas redes sociais *online X e Instagram* e os resultados dos questionários de percepção de suporte social e temperamento.

Sobre o LeXPAPC, foi visto que, comparando-o com Vader e SentiStrength, os resultados foram superiores, apesar de a abrangência ter sido inferior na estratégia de lexicografia para classificação final, dado que os métodos comparados classificam todas as entradas. Contudo, ao aplicar estratégias de funil e votação, em especial o funil, houve uma queda na acurácia, mas as demais medidas de avaliação se tornaram mais consistentes.

Na análise de contrastes observam-se indícios que sugerem uma relação entre o comportamento dos usuários nas plataformas *X* e *Instagram* e seus respectivos temperamentos e percepção de suporte social. As hipóteses foram verificadas por meio das análises realizadas neste trabalho, onde o uso do LeXPAPC para classificar a polaridade das postagens permitiu identificar padrões que apoiam a existência de tais conexões. Estes indícios observados reforçam a ideia de que as expressões online, em termos de sentimentos positivos ou negativos, podem refletir traços subjacentes do temperamento dos indivíduos e sua percepção sobre o suporte social que recebem.

Apesar das dificuldades encontradas na coleta de dados, o que resultou em um baixo volume de dados para analisar, a proposta deste trabalho se apresentou como um ferramental que pode ser aplicado em novas coletas de dados e os resultados experimentais da abordagem para contrastar o real e o *online* mostrou potencial para ser adotado como ferramenta para a área da psiquiatria.

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 criadas para representar o usuário do *X* e do *Instagram*.

7. Limitações e Fragilidades para Validação da Pesquisa

As principais limitações deste estudo decorrem da coleta de dados limitada aos alunos da UFU, refletindo comportamentos e percepções que podem não ser representativos da diversidade mais ampla dos usuários de mídias sociais *online*. Tal especificidade da amostra restringe a generalização dos resultados além do contexto acadêmico.

Adicionalmente, as postagens analisadas representam instantâneos temporais que podem não capturar mudanças longitudinais no comportamento ou percepções dos usuários, especialmente diante de eventos significativos ou tendências culturais. Investigações futuras poderiam focar em analisar como essas dinâmicas afetam as relações estudadas.

Por fim, questões éticas ligadas à privacidade dos dados e ao consentimento dos usuários impõem restrições significativas à replicabilidade direta deste trabalho com o mesmo conjunto de dados. No entanto, os métodos desenvolvidos aqui podem ser adaptados para novas investigações, caso respeitem os princípios éticos para a coleta e análise de dados pessoais *online*.

Referências

- [Brum and Volpe Nunes 2018] Brum, H. and Volpe Nunes, M. d. G. (2018). Building a sentiment corpus of tweets in Brazilian Portuguese. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- [Burth et al. 2019] 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 and Blackwell 1998] 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.
- [de Alcantara et al. 2023] de Alcantara, C., de Faria Paiva, E. R., Barioni, M. C. N., and Junior, L. C. O. (2023). Método para contrastar percepção de suporte social com dados extraídos do twitter em português. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*, pages 31–42. SBC.
- [Dembńska-Krajewska and Rybakowski 2014] Dembińska-Krajewska, D. and Rybakowski, J. (2014). The temperament evaluation of memphis, pisa and san diego autoquestionnaire (temps-a)—an important tool to study affective temperaments. *Psychiatr Pol*, 48(2):261–276.
- [Freitas 2013] 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 and Berah 1996] 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 et al. 2022] 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 and Gilbert 2014] 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.

- [Karam et al. 2005] Karam, E. G., Mneimneh, Z., Salamoun, M., Akiskal, K. K., and Akiskal, H. S. (2005). Psychometric properties of the lebanese–arabic temps-a: a national epidemiologic study. *Journal of affective disorders*, 87(2-3):169–183.
- [Lima and de Castro 2019] 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 et al. 2021] 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 et al. 2022] 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 and Kretzschmar 2008] 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.
- [Myers 1962] Myers, I. B. (1962). The myers-briggs type indicator: Manual (1962).
- [Plank and Hovy 2015] 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 1999] Ribeiro, J. L. P. (1999). Escala de satisfação com o suporte social (esss).
- [Rothbart and Derryberry 1981] Rothbart, M. K. and Derryberry, D. (1981). Development of individual differences in temperament. In Lamb, M. E. and Brown, A. L., editors, *Advances in developmental psychology*, volume 1, pages 37–86. Lawrence Erlbaum Associates.
- [Silva et al. 2019] Silva, E. d., Malheiros, Y., Nunes, R., Antunes, I., and Rêgo, T. (2019). Um conjunto de dados extraído do twitter para análise de sentimentos na língua portuguesa. *Trabalho de Conclusão de Curso em Sistemas de Informação da Universidade Federal da Paraíba*.
- [Thelwall et al. 2010] 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.
- [Thoits 2011] Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior*, 52(2):145–161.
- [Vioules et al. 2018] Vioules, M. J., Moulahi, B., Azé, J., and Bringay, S. (2018). Detection of suicide-related posts in twitter data streams. *IBM Journal of Research and Development*, 62(1):7–1.

[Woodruff et al. 2011] Woodruff, E., Genaro, L. T., Landeira-Fernandez, J., Cheniaux, E., Laks, J., Jean-Louis, G., Nardi, A. E., Versiani, M. C., Akiskal, H. S., and Mendlowicz, M. V. (2011). Validation of the brazilian brief version of the temperament auto-questionnaire temps-a: the brief temps-rio de janeiro. *Journal of affective disorders*, 134(1-3):65–76.