

Evolução Emocional e Comportamental no Reddit: Uma Análise Temporal do Subreddit 'Desabafos'

1st Luiza Viana Afonso

Departamento de Ciéncia da Computaçāo

Universidade Federal de Minas Gerais

Belo Horizonte, Brasil

luiza13lv@gmail.com

Abstract—This study extends the analysis of the subreddit Desabafos by focusing on the temporal dynamics, predictive modeling, and the relationship between sentiment and user engagement in a large-scale Brazilian Portuguese online community. Using weekly and monthly time series derived from posts and comments, the research investigates the evolution of content volume, engagement patterns, and emotional distributions across the main flairs “Desabafo,” “Relacionamento,” and “Dia a Dia.” Trend and seasonal decomposition techniques were applied to identify long-term growth patterns and recurring seasonal behaviors. Univariate forecasting models, including ARIMA and Prophet, were employed to predict posting volume, engagement, and sentiment proportions, with Prophet achieving superior predictive performance. Additionally, multivariate time series models were used to evaluate the influence of sentiment on engagement. The results indicate strong temporal trends and seasonal effects, moderate predictive performance for engagement, higher accuracy in sentiment forecasting, and limited but context-dependent influence of sentiment on user engagement. These findings contribute to a deeper understanding of collective emotional expression and interaction dynamics in anonymous online environments.

Index Terms—Time Series Analysis, Sentiment Analysis, Engagement Prediction, Subreddits, Brazilian Portuguese, Reddit Communities

I. INTRODUÇÃO

As comunidades virtuais de desabafo têm se consolidado como espaços de intensa troca emocional e expressão pessoal. O Reddit, por exemplo, é uma das maiores plataformas de fóruns da internet, com cerca de 1,5 bilhão de usuários registrados e 430 milhões de usuários ativos mensalmente (Wise, 2023). Essa rede social organiza discussões por meio de comunidades temáticas chamadas subreddits, nas quais os usuários interagem de forma majoritariamente anônima por meio de postagens e comentários. Entre esses espaços, destaca-se o subreddit "Desabafo"¹, uma comunidade de língua portuguesa voltada ao compartilhamento de experiências pessoais, desafios emocionais e opiniões sobre a vida cotidiana. Seu caráter anônimo e acolhedor o torna um canal significativo de suporte emocional. Com mais de 369.000 posts, o subreddit é estruturado de modo a facilitar a categorização e a análise de conteúdo por meio do uso obrigatório de flairs, que funcionam como etiquetas que indicam o tema ou o tom

da postagem. As dez flairs disponíveis tais como “Relacionamento”, “Depressão”, “Dia a Dia” e diferentes variantes de “Desabafo”, fornecem pistas importantes sobre o conteúdo e o estado emocional dos autores. Flairs como “Desabafo - sem conselhos” e “Desabafo sem comentário”, por exemplo, indicam a intenção do autor de apenas expressar-se, sem necessariamente buscar interação, evidenciando a diversidade de motivações presentes na comunidade. Diante da relevância e crescimento de plataformas de suporte emocional online, o qual é um fenômeno alinhado ao aumento global do uso de redes sociais, que já ultrapassam 4 bilhões de usuários segundo a Statista [4], torna-se cada vez mais necessário compreender como os sentimentos e os temas abordados influenciam o comportamento dos participantes. Neste contexto, a análise de séries temporais se mostra uma abordagem promissora, especialmente por permitir a identificação de padrões e tendências no comportamento dos usuários ao longo do tempo. De acordo com The Concise Encyclopedia of Statistics [13], uma série temporal pode ser decomposta em quatro componentes principais: tendência de longo prazo, variações sazonais, flutuações cíclicas e variações irregulares. Essa decomposição facilita a compreensão de oscilações e mudanças nos dados, possibilitando, por exemplo, o monitoramento da evolução emocional ou do engajamento dos usuários em comunidades virtuais. Na primeira etapa deste trabalho, foi realizada uma análise quantitativa dos sentimentos e do engajamento dos usuários deste subreddit, revelando uma predominância de sentimentos negativos nas postagens e uma correlação direta entre esses sentimentos e altos níveis de interação. A flair “Desabafo”, por exemplo, concentrou o maior volume de engajamento em posts com sentimento negativo, reforçando o papel das emoções intensas na dinâmica participativa do subreddit. Este estudo aprofunda a caracterização inicial desenvolvida no trabalho realizado na disciplina MSI I e tem como objetivo ampliar a análise do subreddit “Desabafo” por meio de abordagens de natureza textual, semântica, temporal e preditiva.

II. TRABALHOS RELACIONADOS

A análise de sentimentos tem se destacado como uma técnica valiosa para captar opiniões em plataformas de mídia social. Ibrahim e Wang (2022) [14] investigaram a dinâmica de sentimentos de consumidores no contexto do varejo online,

¹<https://www.reddit.com/r/desabafos/>

analisando tweets relacionados a cinco grandes marcas do Reino Unido durante períodos de alta demanda, como Black Friday, Natal e Ano Novo. Por meio da combinação de análise temporal, análise de sentimentos e modelagem de tópicos, os autores identificaram pontos críticos em que as tendências de sentimento sofreram variações significativas. A investigação desses instantes de tempo permitiu compreender quais eventos ou características específicas impulsionaram mudanças nos sentimentos expressos pelos consumidores. Esse estudo destaca a utilidade de métodos de análise textual para mapear flutuações emocionais ao longo do tempo em plataformas digitais e oferece contribuições relevantes para a compreensão do comportamento do usuário, especialmente em contextos de grande volume de interações. A abordagem adotada pelos autores reforça o valor da análise integrada de sentimentos e tópicos como forma de revelar padrões latentes em comunidades online — uma perspectiva que também se mostra pertinente na análise de plataformas de apoio emocional como o subreddit “Desabafo”.

Além disso, Zhang et al. (2022) [15] investigaram o engajamento de adolescentes em redes sociais durante a pandemia da COVID-19, com foco na frequência de postagens e na expressão emocional em subreddits voltados para adolescentes e saúde mental. Por meio da análise temporal de dados semanais ao longo de 2020, os autores observaram que eventos significativos da pandemia influenciaram diretamente o comportamento dos usuários, tanto em termos de atividade quanto na manifestação de emoções negativas. Adolescentes que participavam de subreddits voltados à saúde mental, como r/Depression, r/Anxiety e r/SuicideWatch, apresentaram maior frequência de postagens e comentários, além de maior intensidade emocional negativa, em comparação àqueles que interagiam apenas no r/Teenagers. A aplicação da técnica de Interrupted Time Series Analysis (ITSA) evidenciou que, após o início da pandemia, houve uma reversão nas tendências de queda para aumento tanto na frequência de postagens quanto na negatividade emocional, especialmente entre adolescentes com histórico de participação em comunidades de saúde mental. O estudo reforça a importância de se analisar padrões emocionais e comportamentais em plataformas como o Reddit, especialmente em contextos de crise, e destaca como esses espaços funcionam como válvulas de escape e interação social — aspecto também observado na dinâmica emocional do subreddit “Desabafo”.

Já Le et al. (2012) [16] propuseram uma abordagem preditiva baseada em séries temporais para analisar a dinâmica coletiva de sentimentos em redes sociais como Twitter e Facebook. O foco do estudo foi prever mudanças no sentimento coletivo sem a necessidade de analisar em detalhe usuários individuais ou estruturas de rede, priorizando uma modelagem estatística agregada. Os autores argumentam que essa análise em larga escala permite que governos e empresas reajam de forma estratégica à opinião pública, principalmente em cenários de sentimentos negativos, por meio de ações como desmentido de boatos ou a publicação de mensagens equilibradas. Os resultados mostraram uma acurácia superior

a 85% na previsão da direção das mudanças de sentimento, evidenciando o potencial dessas ferramentas preditivas. A perspectiva apresentada neste trabalho dialoga com a proposta deste estudo, que também busca compreender as variações nos sentimentos ao longo do tempo em comunidades online, como o subreddit “Desabafo”, além de explorar como essas emoções coletivas se associam ao engajamento dos usuários.

Complementando essa abordagem, Silveira, Couto da Silva e Murai (2020) [17] investigaram o efeito das interações em comunidades de saúde mental no Reddit sobre a variação do tom emocional (TE) dos usuários ao longo do tempo. O estudo revelou que, em 68% dos casos, o TE dos usuários apresentou uma mudança positiva após interações na plataforma, sugerindo que participar dessas comunidades pode contribuir para uma melhora subjetiva no estado emocional. Além disso, os autores propuseram modelos preditivos capazes de estimar essa variação com boa acurácia, permitindo o monitoramento contínuo dos usuários. Embora o estudo reconheça a limitação de não poder confirmar se essas mudanças refletem uma melhora clínica real, seus achados reforçam o potencial das redes sociais como ferramentas de apoio emocional. Essa abordagem é particularmente relevante para comunidades como o subreddit Desabafo, onde usuários buscam acolhimento e expressão emocional em momentos de vulnerabilidade.

III. METODOLOGIA

Este projeto foi conduzido em etapas voltadas à análise temporal das postagens e comentários do subreddit Desabafo, com foco na identificação de padrões ao longo do tempo, na decomposição das séries, na previsão de engajamento e na modelagem multivariada entre sentimentos e interação dos usuários. As etapas metodológicas são descritas a seguir.²

A. Organização das Séries Temporais

As séries temporais foram construídas a partir das postagens e comentários previamente processados e classificados quanto ao sentimento. Inicialmente, os dados foram organizados em diferentes resoluções temporais, com foco principal na agregação semanal, de modo a capturar padrões de curto e médio prazo sem comprometer a estabilidade estatística das séries. Para cada flair analisada (Desabafo, Relacionamento e Dia a Dia), foram derivadas três séries principais: volume de postagens, volume de comentários e engajamento, medido pela soma do número de upvotes e pelo número de comentários por post. Além disso, foram construídas séries temporais representando a proporção semanal de sentimentos (negativo, neutro e positivo), obtidas a partir da razão entre a quantidade de textos classificados em cada sentimento e o total de textos da respectiva semana. Esse procedimento permitiu analisar não apenas a evolução do volume de atividade, mas também a dinâmica emocional da comunidade ao longo do tempo.

²Para auxílio na revisão de texto deste trabalho, foi utilizada a ferramenta ChatGPT (OpenAI).

B. Decomposição das Séries Temporais

Com o objetivo de identificar padrões estruturais nas séries, foi aplicada a técnica de decomposição sazonal aditiva, que separa cada série em três componentes principais: tendência, sazonalidade e ruído. A tendência representa a evolução de longo prazo da atividade no subreddit, enquanto a sazonalidade evidencia ciclos recorrentes associados ao calendário anual. O termo residual captura variações aleatórias não explicadas pelos demais componentes. A decomposição foi realizada individualmente para cada flair e para as séries de postagens e comentários. A partir desses componentes, foi possível comparar os padrões sazonais médios entre as flairs, permitindo identificar meses com maior e menor atividade relativa. Essa análise é fundamental para compreender como fatores temporais externos, como períodos letivos, férias e datas comemorativas, influenciam o comportamento dos usuários.

C. Análise Comparativa Temporal entre Flairs

A partir das componentes sazonais extraídas na etapa de decomposição, foi realizada uma análise comparativa entre as flairs quanto aos seus ciclos anuais médios. Para cada mês do ano, foi calculada a média da componente sazonal, permitindo identificar períodos de maior e menor atividade relativa. Essa análise possibilitou compreender como o comportamento dos usuários varia ao longo do ano e como diferentes tipos de conteúdo (emocional, relacional e cotidiano) respondem de maneira distinta aos ciclos temporais.

D. Análise de Séries Temporais e Predição do Volume, Sentimento e Engajamento

A etapa de modelagem preditiva foi conduzida a partir da construção de séries temporais semanais para as variáveis de interesse: volume de postagens por flair, engajamento dos posts (número médio de comentários e número médio de upvotes por post), engajamento dos comentários (número médio de upvotes por comentário) e proporções dos sentimentos (negativo, neutro e positivo).

Inicialmente, foram realizados testes com agregação mensal; contudo, verificou-se que a quantidade de observações era insuficiente para o ajuste estável dos modelos preditivos. Dessa forma, optou-se pela agregação semanal, o que permitiu maior robustez estatística e melhor desempenho dos modelos.

Para a predição do volume de postagens, do engajamento e das proporções de sentimentos, foram utilizados modelos probabilísticos de séries temporais, com destaque para o Prophet (2017) [18], desenvolvido pela Meta (antigo Facebook). Esse modelo foi escolhido por sua capacidade de modelar, de forma simultânea, componentes de tendência, sazonalidade e possíveis variações não lineares, além de sua robustez frente a séries ruidosas e com quebras estruturais.

Os dados foram divididos em conjuntos de treino e teste segundo uma abordagem temporal, respeitando a ordem cronológica das observações. Os modelos foram ajustados exclusivamente sobre o conjunto de treino, sendo as previsões realizadas para o período correspondente ao conjunto de teste. Esse procedimento foi aplicado de maneira análoga

às séries de volume de postagens, comentários, engajamento e sentimentos, possibilitando a comparação entre a previsibilidade dessas diferentes dimensões do comportamento da comunidade.

E. Modelagem Multivariada entre Sentimento e Engajamento

Para investigar a existência de relações temporais entre a variação emocional das postagens e o engajamento gerado, foi adotada uma abordagem de modelagem multivariada. Nesse contexto, o engajamento foi tratado como variável dependente, enquanto as proporções dos sentimentos negativo, neutro e positivo foram utilizadas como variáveis exógenas. Foi empregado o modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables), que permite incorporar fatores externos à dinâmica da série principal. O ajuste do modelo foi realizado a partir do conjunto de treino, e as previsões foram geradas para o período de teste, utilizando as observações reais das variáveis exógenas. A significância estatística dos coeficientes foi avaliada por meio dos valores de p-value, permitindo verificar se as variações nos sentimentos exercem influência estatisticamente relevante sobre o engajamento.

F. Avaliação dos Modelos Preditivos

A avaliação do desempenho preditivo foi realizada por meio das métricas MAE (Erro Absoluto Médio) e RMSE (Raiz do Erro Quadrático Médio). O MAE fornece uma medida de erro diretamente interpretável nas mesmas unidades da variável predita, enquanto o RMSE penaliza erros de maior magnitude de forma mais acentuada, sendo mais sensível a valores extremos. O uso conjunto dessas métricas permite mensurar a precisão dos modelos ao comparar valores observados e preditos. Essas métricas foram aplicadas tanto às previsões univariadas quanto às multivariadas. Além das métricas numéricas, foram analisados graficamente os ajustes dos modelos, por meio da comparação entre séries reais e preditas, possibilitando a inspeção visual do comportamento das previsões ao longo do tempo. Essa análise visual foi fundamental para verificar a capacidade dos modelos em acompanhar variações abruptas, mudanças de regime e padrões sazonais.

IV. RESULTADOS

Nesta seção são apresentados os resultados obtidos a partir da análise temporal das séries de postagens, comentários, sentimentos e engajamento no subreddit Desabafo. São discutidos os padrões de tendência e sazonalidade, bem como o desempenho dos modelos preditivos e a relação entre sentimento e engajamento.

A. Análise textual e semântica

A análise do conteúdo textual e semântico dos posts e comentários revela padrões discursivos distintos e uma evolução na complexidade da linguagem utilizada ao longo do tempo nas diferentes flairs (categorias).

a) *Evolução da Diversidade Lexical (TTR)*: A diversidade lexical, medida pelo Type-Token Ratio (TTR), indica a riqueza do vocabulário e a variação da linguagem empregada. A figura 1 representa o gráfico de evolução do TTR (média móvel de 3 meses) demonstra uma tendência de queda generalizada em todos os flairs desde o início de 2020. A flair "Dia a Dia" e "Relacionamento" iniciam a série com a maior diversidade lexical, ambos acima de 0.5 em 2020-02, mas seguem uma declinação constante e acentuada até meados de 2022. A flair "Desabafo" sempre se manteve com a menor diversidade, partindo de cerca de 0.4 e atingindo um patamar de estabilidade extremamente baixo (próximo de 0.15) a partir de 2022-03. A partir de 2022, todos os flairs entram em um período de baixa diversidade lexical, sugerindo que, com o amadurecimento e o aumento de volume da comunidade, os usuários passaram a empregar um vocabulário mais limitado, repetitivo ou específico da comunidade, caracterizando uma linguagem mais concisa e padronizada.

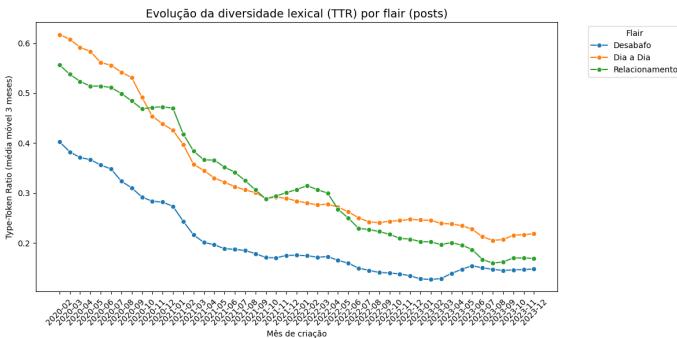


Fig. 1. Evolução TTR por Flair (posts)

b) *Comprimento dos Textos*: A análise do comprimento dos textos (posts e comentários) oferece informações sobre a profundidade e o esforço dedicado às interações. Entre 2020 e 2022, o comprimento médio dos posts era relativamente estável e similar entre os três flairs (entre 50 e 70 caracteres/palavras). No segundo semestre de 2023, ocorreu um crescimento abrupto. As flairs "Desabafo" e "Relacionamento" passaram a apresentar posts significativamente mais longos (próximos ou acima de 90), indicando que os usuários estão dedicando mais espaço para expressar suas emoções e relatar contextos complexos nessas categorias. A flair "Dia a Dia" também cresceu, mas manteve-se consistentemente com os posts de menor comprimento. O comprimento médio dos comentários apresenta uma tendência geral de queda em todas as flairs ao longo do tempo. Relacionamento e Desabafo possuem os comentários mais longos, indicando que as interações nessas áreas (que demandam conselhos e apoio) são mais detalhadas ou reflexivas. A flair Dia a Dia registra o menor comprimento médio de comentários, sugerindo interações mais breves, diretas e menos elaboradas.

c) *Frequência de Palavras Chave por Mês*: A análise da frequência das palavras-chave na figura 2 das três flairs por meio dos heatmaps (mapas de calor) revela como o foco e o tom do discurso evoluíram e se diferenciam em cada categoria

da comunidade. Na flair Desabafo, há uma clara intensificação na frequência de diversas palavras a partir de meados de 2022 (coincidindo com a queda na Diversidade Lexical), onde as células se tornam de azul escuro (alta frequência). Palavras como "mim", "sempre", "sei", "vida" e "anos" são proeminentes e se intensificam, sugerindo que os desabafofes frequentemente envolvem a autoavaliação de problemas de longa data ou padrões de comportamento recorrentes. Termos como "alguém" e "coisas" também são frequentes, indicando a busca por validação ou ajuda, e a necessidade de incluir diversos elementos para contextualizar o desabafo. O uso de palavras diretamente relacionadas ao tema na flair Relacionamento, como "relacionamento" e "namoro", é consistentemente alto no período final, validando o foco da categoria. Palavras como "sinto", "sei", "falar", "quereria" e "pessoas" são intensas (azul escuro) no período de 2022-2023. Isso sugere que a discussão é altamente emocional e deliberativa, focada na expressão de sentimentos, nas dificuldades de comunicação e na busca por compreensão das outras partes. O aumento na frequência de palavras como "alguém" e "gente" indica que a comunidade é usada para buscar aconselhamento, opiniões coletivas ou experiências compartilhadas sobre as dinâmicas de relacionamento.

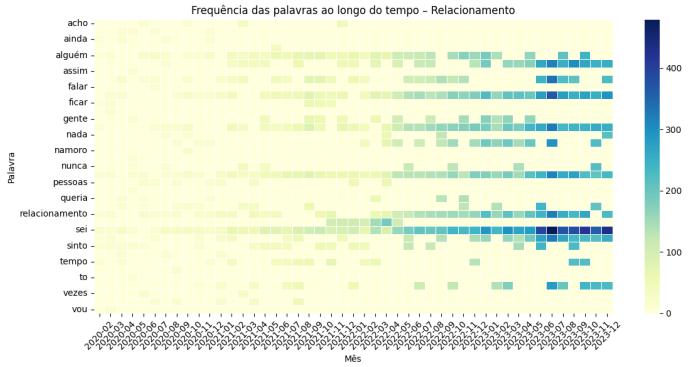


Fig. 2. Frequência Das Palavras - Flair Relacionamento

Na flair Dia a Dia, palavras como "fazer", "jogo", "casa" e "coisas" são as mais frequentes. Isso reflete a natureza descritiva da categoria, que trata de atividades práticas, hobbies e o relato de eventos corriqueiros da vida. O padrão de alta frequência ("fazer", "pessoas", "sempre") se consolida a partir de 2022, de forma semelhante aos outros flairs, mas com uma paleta de cores ligeiramente menos escura que o Desabafo, sugerindo que, embora o volume tenha aumentado, a concentração de tópicos é mais dispersa. A tendência geral observada é que, com o amadurecimento e a consolidação do tráfego na comunidade, o vocabulário se tornou mais concentrado e a frequência das palavras-chave se intensificou nos três flairs a partir de 2022, mesmo com a queda da diversidade lexical (TTR).

B. Evolução Temporal do Volume de Postagens

A análise da evolução temporal do volume de postagens no subreddit r/Desabafo, indicada pela figura 3 evidencia

um crescimento expressivo da atividade ao longo da série histórica, especialmente a partir de 2020. Considerando as três flairs analisadas (Desabafo, Dia a Dia e Relacionamento) observamos que todos apresentam tendência de alta, embora em magnitudes distintas.

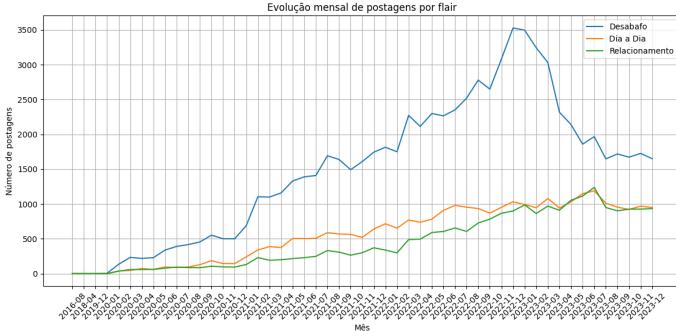


Fig. 3. Evolução Mensal de Posts por Flair

A flair Desabafo concentra, de forma consistente, o maior volume absoluto de postagens em todo o período analisado. Entre 2016 e 2019, o número mensal de postagens permaneceu relativamente baixo e estável. A partir de 2020, ocorre uma mudança estrutural clara no padrão da série, com crescimento acelerado em 2021 e 2022, culminando em picos superiores a 3.500 postagens mensais em 2023. Após esse pico, é possível notar uma leve redução em 2024, ainda que em patamares significativamente superiores aos anos iniciais. Esse comportamento sugere um processo de consolidação do subreddit como espaço de expressão emocional ao longo do período pandêmico e pós-pandêmico. O crescimento estrutural a partir de 2020 também influencia a interpretação visual da série ao longo de todo o período. Nos anos iniciais, as variações mensais existem, porém aparecem visualmente atenuadas devido à grande diferença de escala em relação ao volume registrado após o grande crescimento de participação da comunidade. A flair Dia a Dia apresenta crescimento mais moderado, porém contínuo. Assim como em Desabafo, há um aumento mais acentuado a partir de 2020, com estabilização em torno de 900 a 1.100 postagens mensais nos anos mais recentes. A trajetória indica que conteúdos relacionados à rotina ganharam maior relevância com o amadurecimento da comunidade. Já o flair Relacionamento apresenta o menor volume absoluto entre os três, mas também exibe tendência crescente. O crescimento torna-se mais evidente a partir de 2021, atingindo valores próximos de 1.200 postagens mensais em 2023, seguido de leve retração no período final da série. Esse comportamento sugere que discussões afetivas também acompanharam o aumento geral de participação no subreddit.

a) *Decomposição Temporal - Flair Desabafo*: A decomposição da série temporal da flair Desabafo (Figura 4) permite identificar, de forma mais precisa, os componentes de tendência, sazonalidade e resíduo. O componente de tendência confirma o crescimento estrutural observado no gráfico bruto, com elevação acentuada a partir de 2020 e pico em torno de 2023, seguido de uma leve inflexão negativa. Embora visualmente

o período anterior a 2020 apresente um comportamento quase linear, essa impressão decorre principalmente da diferença de grandeza entre o volume de postagens nos primeiros anos e o volume observado após o aumento expressivo da participação na comunidade. As variações existentes no início da série tornam-se menos perceptíveis no gráfico agregado devido a esse efeito de escala. A sazonalidade mensal apresenta um padrão claramente recorrente, com variações sistemáticas ao longo do ano. Observam-se meses com contribuição positiva consistente para o volume de postagens, enquanto outros apresentam quedas regulares, indicando a influência de ciclos anuais no comportamento dos usuários. O componente de resíduo revela maior volatilidade a partir de 2021, com picos positivos e negativos mais intensos, sugerindo a ocorrência de eventos externos ou de dinâmicas específicas da comunidade que provocam desvios temporários em relação ao padrão esperado.

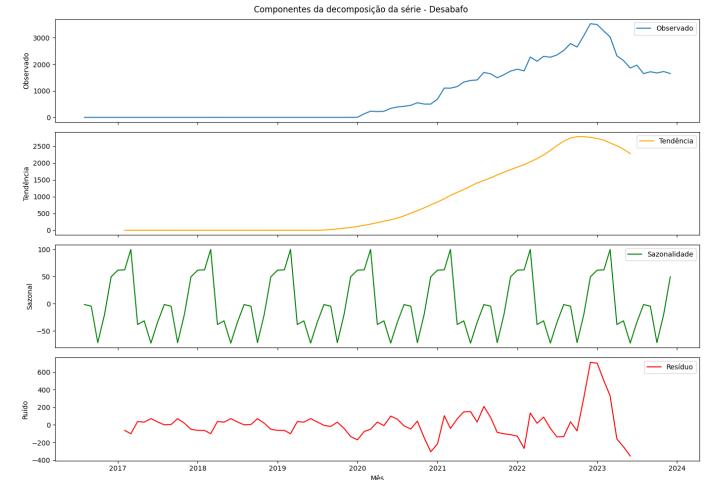


Fig. 4. Componentes da Decomposição da Série (Flair Desabafo)

b) *Decomposição Temporal - Flair Dia a Dia*: O componente de tendência da série temporal da flair Dia a Dia representado na figura 5 evidencia um crescimento lento e gradual nos anos iniciais, seguido por uma aceleração acentuada a partir de 2020, atingindo um patamar elevado entre o final de 2022 e o início de 2023. Esse comportamento indica uma elevação estrutural sustentada no volume de postagens, sugerindo a consolidação da popularidade dessa categoria ao longo do período analisado. A sazonalidade mensal apresenta um padrão claramente recorrente e estável ao longo dos anos, com variações cíclicas que se repetem anualmente, indicando a influência persistente de ciclos temporais no comportamento dos usuários. O componente de resíduo, por sua vez, mantém flutuações relativamente estáveis nos primeiros anos, mas passa a apresentar maior volatilidade a partir de 2021, com picos positivos mais intensos em 2022 e 2023, o que sugere a atuação de eventos externos como campanhas virais, picos de engajamento em datas especiais ou mudanças nas regras de moderação da comunidade, os quais podem provocar desvios temporários em relação ao padrão esperado.

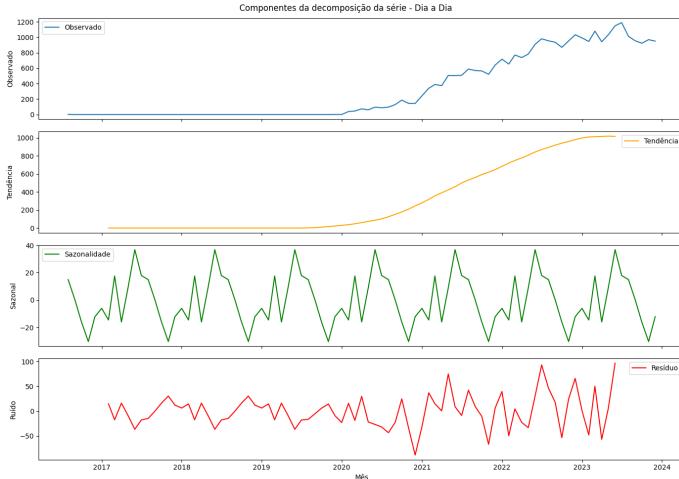


Fig. 5. Componentes da Decomposição da Série (Flair Dia a Dia)

c) Decomposição Temporal - Flair Relacionamento:

Para a flair Relacionamento, a decomposição revela um padrão de crescimento mais suave e progressivo, como demonstrado na figura 6. A tendência é claramente ascendente, com aceleração a partir de 2021. A sazonalidade é menos intensa que nos demais flairs, indicando menor influência de ciclos anuais sobre esse tipo de conteúdo. O componente residual apresenta menor variância quando comparado ao flair Desabafo, o que sugere que as discussões sobre relacionamento tendem a seguir um comportamento mais estável ao longo do tempo, com menos picos abruptos.

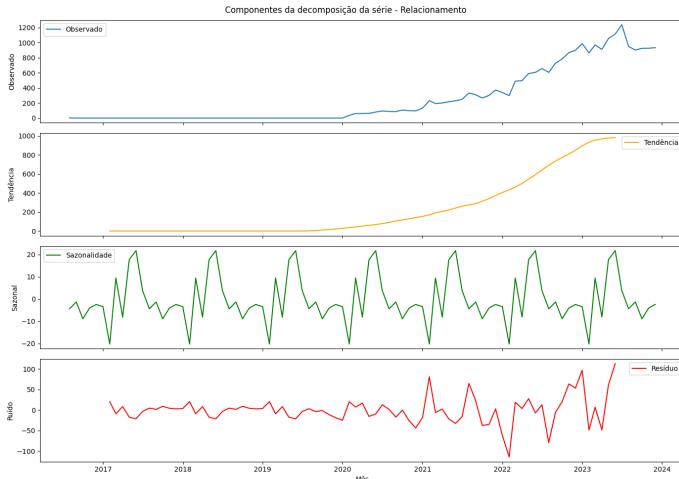


Fig. 6. Componentes da Decomposição da Série (Flair Relacionamento)

C. Análise da Sazonalidade por Flair

A comparação da sazonalidade média mensal na Figura 7 entre as flairs Desabafo, Relacionamento e Dia a Dia revelou comportamentos distintos ao longo do ano. A flair Desabafo apresentou maior intensidade sazonal positiva nos meses de março e dezembro, enquanto registrou quedas expressivas

nos meses de abril, junho e outubro. Esses padrões indicam que os desabafo tendem a se intensificar em períodos que coincidem com transições de ciclos sociais, como início e fim de ano letivo. A flair Relacionamento apresentou variações mais suaves ao longo do ano, com picos moderados entre maio e junho, sugerindo uma maior estabilidade temática. Já a flair Dia a Dia apresentou seu maior pico sazonal no mês de junho, com quedas nos últimos meses do ano, indicando maior concentração de relatos cotidianos no período intermediário do calendário. Essas diferenças indicam que cada flair responde de maneira distinta aos ciclos temporais, refletindo padrões específicos de comportamento e de expressão emocional dos usuários.

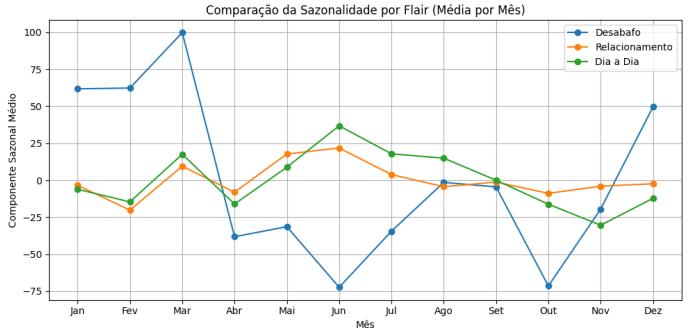


Fig. 7. Sazonalidade por Flair

V. PREDIÇÃO DO VOLUME, SENTIMENTO E ENGAJAMENTO

Em uma primeira etapa, as tentativas de predição foram realizadas com séries mensais de volume de postagens, sentimentos e engajamento. No entanto, observou-se que a quantidade de pontos temporais disponíveis nesse nível de agregação era relativamente pequena, o que comprometeu o ajuste adequado dos modelos e a estabilidade das previsões. Diante dessa limitação, optou-se por realizar a agregação das séries para a escala semanal, aumentando o número de observações e, consequentemente, a robustez dos modelos preditivos. Para a tarefa de previsão, foram testados modelos da família ARIMA/SARIMA e o modelo Prophet, desenvolvido pelo Facebook, amplamente utilizado em séries temporais com tendência e sazonalidade complexas. De forma consistente, o modelo Prophet apresentou desempenho superior ao ARIMA na maioria dos experimentos, especialmente na presença de padrões sazonais não estritamente lineares e de variações abruptas ao longo do tempo. Assim, os resultados apresentados nesta seção referem-se majoritariamente aos modelos ajustados com Prophet. No que se refere à previsão univariada do volume de postagens, os erros obtidos foram relativamente elevados, refletindo a alta variabilidade das séries ao longo do tempo. Como exemplo, na previsão do volume semanal de postagens da flair Dia a Dia, o modelo Prophet (com sazonalidade anual) obteve MAE de 21,93 e RMSE de 26,31. Esses valores indicam que, embora o modelo consiga capturar a tendência geral e parte da sazonalidade, ainda há dificuldades

em prever com precisão oscilações mais bruscas no volume de publicações, especialmente em períodos de crescimento acelerado ou queda abrupta. Em contraste, quando aplicados à predição da proporção dos sentimentos, os modelos apresentaram desempenho substancialmente melhor. Um exemplo representativo é a previsão da proporção de sentimento negativo na flair Desabafo, na qual o modelo Prophet alcançou MAE de 0,0470 e RMSE de 0,0555. Esses resultados indicam alto grau de precisão, o que pode ser explicado pelo fato de que as séries de proporção de sentimentos apresentam variações mais suaves e estruturadas ao longo do tempo, tornando o processo de modelagem mais estável do que aquele aplicado ao volume bruto de postagens. Também foram realizadas previsões para as métricas de engajamento, em especial o número de upvotes em posts. Como exemplo, na previsão da média semanal de upvotes dos posts da flair Relacionamento, o modelo obteve MAE de 5,08 e RMSE de 5,92, sugerindo desempenho satisfatório na captura do comportamento médio de engajamento dessa categoria. Ainda assim, eventos pontuais de grande repercussão continuam sendo difíceis de antecipar com precisão.

VI. RELAÇÃO ENTRE SENTIMENTO E ENGAJAMENTO: MODELO MULTIVARIADO

Com o objetivo de investigar se a distribuição dos sentimentos nos posts exerce influência sobre o engajamento dos usuários, foram ajustados modelos multivariados do tipo SARIMAX, considerando como variável dependente as métricas de engajamento (número de comentários e upvotes) e, como variáveis exógenas, as proporções mensais dos sentimentos negativo, neutro e positivo. A análise foi realizada separadamente para cada flair.

De modo geral, os resultados indicam que, na maior parte das flairs analisadas, os coeficientes associados aos sentimentos não apresentaram significância estatística, com p-values elevados, o que sugere ausência de evidência robusta de que as variações na distribuição de sentimentos expliquem diretamente o engajamento nessas categorias. Esse comportamento aponta que, em geral, o engajamento parece ser mais fortemente explicado por fatores estruturais da série (tendência, sazonalidade e dependência temporal) do que pelas variações emocionais agregadas. Entretanto, no caso específico da flair Desabafo, ao considerar o número de comentários como variável de engajamento, observou-se um resultado estatisticamente relevante para o sentimento negativo, conforme ilustrado na Figura 8 de coeficientes do modelo.

O coeficiente estimado para o sentimento negativo foi positivo ($\beta = 25,11$) e estatisticamente significativo ao nível de 5% ($p = 0,037$), indicando que aumentos na proporção de posts classificados como negativos estão associados a um aumento no número médio de comentários. Este resultado sugere que conteúdos com carga emocional negativa tendem a gerar maior interação na comunidade, possivelmente por despertarem mais respostas de apoio, aconselhamento ou identificação por outros usuários. Por outro lado, os sentimentos neutro ($p = 0,853$) e positivo ($p = 0,235$) não apresentaram significância estatística,

	coef	std err	z	P> z	[0.025	0.975]
negativo	25.1055	13.706	1.832	0.037	-1.758	51.969
neutro	-3.6279	19.640	-0.185	0.853	-42.122	34.866
positivo	-21.4776	18.085	-1.188	0.235	-56.923	13.968
ar.L1	-0.6791	0.149	-4.543	0.000	-0.972	-0.386
ma.L1	0.9997	78.110	0.013	0.990	-152.093	154.092
ar.S.L12	0.2034	0.144	1.408	0.159	-0.080	0.487
ma.S.L12	0.1176	0.529	0.223	0.824	-0.918	1.154
sigma2	2.2776	177.191	0.013	0.990	-345.010	349.565

Fig. 8. Resumo dos Coeficientes do Modelo

indicando que, isoladamente, essas categorias emocionais não possuem poder explicativo relevante sobre o volume de comentários no flair Desabafo. Assim, o efeito observado parece estar concentrado especificamente na dimensão negativa do conteúdo emocional. Os termos autorregressivos do modelo reforçam a forte dependência temporal da série, com destaque para o componente AR(1), que apresentou coeficiente significativo ($p < 0,001$). Esse resultado confirma que o número de comentários em um dado mês depende fortemente do comportamento observado no mês anterior, o que é típico em séries de engajamento em redes sociais. Já os termos de média móvel e sazonais não apresentaram significância estatística. As métricas de erro do modelo para o número de comentários na flair Desabafo foram MAE = 4,23 e RMSE = 5,18, indicando bom desempenho preditivo, com desvios médios relativamente baixos em relação à escala da variável analisada. Além disso, os testes de diagnóstico dos resíduos (Ljung-Box, Jarque-Bera e teste de heterocedasticidade) não indicaram violações severas dos pressupostos do modelo, sugerindo ajuste adequado.

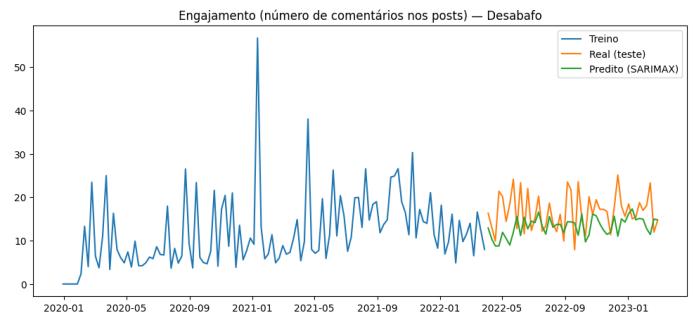


Fig. 9. Predição de Engajamento na Flair Desabafo

VII. CONCLUSÃO

Este trabalho teve como objetivo aprofundar a análise temporal, preditiva e relacional dos dados do subreddit Desabafo, dando continuidade às investigações realizadas na primeira etapa do MSI. A partir da modelagem de séries temporais, foi possível compreender, de forma mais detalhada, a evolução do volume de postagens, do engajamento e dos sentimentos ao longo do tempo, bem como avaliar a capacidade de predição desses fenômenos em um ambiente social altamente dinâmico e sujeito a variações abruptas. A análise da evolução temporal evidenciou um crescimento expressivo da atividade no subreddit a partir de 2020, seguido por um período de estabilização

e leve retração nos anos mais recentes. As decomposições das séries mostraram que, embora exista uma tendência de crescimento bem definida para as flairs analisadas, a sazonalidade apresenta padrões recorrentes ao longo do ano, com variações específicas para cada categoria temática. Além disso, a presença de resíduos com alta variabilidade reforça o caráter imprevisível do comportamento dos usuários, especialmente em períodos marcados por eventos externos de grande impacto. No âmbito da predição, verificamos que a agregação semanal dos dados foi fundamental para a obtenção de modelos mais precisos, uma vez que a granularidade mensal apresentou quantidade insuficiente de observações. O modelo Prophet demonstrou desempenho superior aos modelos ARIMA na maioria dos cenários, especialmente na captura de tendências e padrões sazonais não lineares. Observou-se que as previsões de volume de postagens apresentaram erros relativamente elevados, refletindo a natureza volátil dessas séries. Em contrapartida, as previsões das proporções de sentimentos alcançaram maior precisão, evidenciando que a dinâmica emocional coletiva apresenta variações mais suaves e melhor estruturadas ao longo do tempo. As previsões de engajamento exibiram desempenho intermediário, com boa representação do comportamento médio, mas limitações na antecipação de picos pontuais de interação. A análise multivariada da relação entre as proporções dos diferentes sentimentos (Positivo, Negativo e Neutro) e o engajamento indicou que, de modo geral, essas variáveis emocionais não apresentaram forte capacidade explicativa sobre o engajamento nas diferentes flairs, uma vez que a maioria dos coeficientes não se mostrou estatisticamente significativa. Neste contexto multivariado, onde múltiplas proporções de sentimento atuaram como variáveis preditoras do engajamento (comentários + upvotes), um caso específico foi observado na flair Desabafo: a proporção de sentimento negativo apresentou indícios de associação com o número de comentários, sugerindo que postagens com maior carga emocional negativa tendem a estimular mais interações. Ainda assim, esse efeito foi limitado e não se generalizou para as demais categorias, nem para as outras proporções de sentimentos, indicando que o engajamento é fortemente influenciado por fatores adicionais, como o conteúdo temático, a identificação dos usuários com o relato e eventos externos à plataforma. De forma geral, os resultados obtidos reforçam a complexidade das dinâmicas sociais em ambientes digitais, nas quais padrões temporais, emocionais e de interação se entrelaçam de maneira não linear. Este trabalho contribui ao demonstrar que a análise conjunta de séries temporais, sentimentos e engajamento é uma abordagem promissora para a compreensão do comportamento coletivo em redes sociais, ao mesmo tempo em que evidencia as limitações dos modelos preditivos tradicionais diante da elevada volatilidade desses dados.

VIII. AGRADECIMENTOS

Gostaria de agradecer à Professora Ana Paula Couto pelo apoio, orientação e contribuição ao longo do desenvolvimento deste trabalho.

REFERENCES

- [1] Silveira, B., da Silva, A. P. C., and Murai, F. Análise de comunidades de suporte a transtornos de saúde mental do reddit. In Anais do VII Brazilian Workshop on Social Network Analysis and Mining. SBC. 2018.
- [2] J. Wise. Reddit users: How many people use reddit in 2023? <https://earthweb.com/how-many-people-use-reddit/>. 2023.
- [3] Post Flair <https://support.reddithelp.com/hc/en-us/articles/15484545678996-Post-Flair>. 2024.
- [4] Statista. 2022. Number of social media users worldwide from 2017 to 2028. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. 2024.
- [5] Lima, L., Silveira, B., da Silva, A. P. C. Toxic Content Detection in Online Social Networks: A New Dataset from Brazilian Reddit Communities. 2024.
- [6] Silveira, B., da Silva, A. P. C., and Murai, F. (2018). Análise de comunidades de suporte a transtornos de saúde mental do reddit. In Anais do VII Brazilian Workshop on Social Network Analysis and Mining. SBC. 2018
- [7] Silveira, B., da Silva, A. P. C., and Murai, F. (2020). Modelos de Previsão do Tom Emocional de Usuários em Comunidades de Saúde Mental no Reddit. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM). 2020.
- [8] Bhandari, A., & Sun, B. An online home for the homeless: A content analysis of the subreddit r/homeless. *New Media & Society*. 2023.
- [9] G. Piorino, V. Moreira, L. Lima, A. Pagano, and A. Silva. "Análise de sentimentos de conteúdo compartilhado em comunidades brasileiras do Reddit: Avaliação de um conjunto de dados rotulados por humanos", in Proceedings of the 30th Brazilian Symposium on Multimedia and the Web, Juiz de Fora/MG, pp. 54-62, doi: <https://doi.org/10.5753/webmedia.2024.242020>. 2024
- [10] Elbagir S, Yang J Twitter sentiment analysis using natural language toolkit and Vader sentiment. In: Proceedings of the international multi-conference of engineers and computer scientists, Vol. 122, pp 16. 2019.
- [11] Ayyalasomayajula, Madan Mohan Tito & Agarwal, Akshay & Khan, Shahnawaz. Reddit social media text analysis for depression prediction: using logistic regression with enhanced term frequency-inverse document frequency features. *International Journal of Electrical and Computer Engineering (IJECE)*. 2024.
- [12] Klaifer Garcia and Lilian Berton. 2021. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing* 101, 107057. <https://doi.org/10.1016/j.asoc.2020.107057>. 2021
- [13] Time Series. In: *The Concise Encyclopedia of Statistics*. Springer, New York, NY. <https://doi.org/10.1007/978-0-387-32833-1-401>. 2008.
- [14] Noor Farizah Ibrahim, Xiaojun Wang, Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media, *Computers in Human Behavior*, Volume 96, 2019, Pages 32-45, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2019.02.004>.
- [15] Zhang, S.; Liu, M.; Li, Y.; Chung, J.E. Teens' Social Media Engagement during the COVID-19 Pandemic: A Time Series Examination of Posting and Emotion on Reddit. *Int. J. Environ. Res. Public Health* 2021, 18, 10079. <https://doi.org/10.3390/ijerph181910079>
- [16] Le T. Nguyen, Pang Wu, William Chan, Wei Peng, and Ying Zhang. 2012. Predicting collective sentiment dynamics from time-series social media. In Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM '12). Association for Computing Machinery, New York, NY, USA, Article 6, 1–8. <https://doi.org/10.1145/2346676.2346682>
- [17] Silveira, B., da Silva, A. P. C., and Murai, F. (2020). Modelos de Previsão do Tom Emocional de Usuários em Comunidades de Saúde Mental no Reddit. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM).
- [18] META. Prophet: Forecasting at scale. Meta Open Source, 2017. Disponível em: <https://facebook.github.io/prophet/>. Acesso em: 10/11/2025.