

Análise e Previsão de Postagens

João Antônio Santos Carvalho

2025

1 Introdução

Este projeto tem como objetivo analisar o engajamento dos posts do perfil oficial do The New York Times na rede social Bluesky. A aplicação interativa, desenvolvida em Streamlit, possibilita a coleta e visualização de dados, além de oferecer análises avançadas, como previsão de engajamento, modelagem de tópicos e análise de sentimentos expressos nos textos. Para isso, utilizamos todos os posts publicados pelo perfil do New York Times como base, proporcionando insights sobre as palavras mais frequentes nas postagens, a variação do engajamento ao longo do tempo e previsões baseadas nas características dos posts.

2 Tecnologias Utilizadas

O projeto é implementado em Python, utilizando diversas ferramentas que incluem Streamlit para a interface, Pandas para manipulação de dados, e PIL para processamento de imagens. Para visualização de dados, empregamos Seaborn e Matplotlib, enquanto o processamento de linguagem natural é realizado com NLTK, VADER e LDA. Além disso, Folium é usado para criar mapas interativos e ARIMA para previsão de séries temporais. A coleta de posts é feita por meio da API do Bluesky, complementada pela geração de WordCloud para visualização dos dados.

3 Funcionalidades

A aplicação permite uma análise detalhada dos posts de um usuário do Bluesky, oferecendo informações essenciais para entender seu padrão de engajamento. O usuário pode selecionar um perfil específico e definir parâmetros, como a quantidade de posts a serem analisados por iteração, o número total de iterações e o idioma predominante nas postagens. Também é possível configurar a quantidade de dias para a previsão do engajamento, proporcionando uma análise mais aprofundada do desempenho futuro das publicações.

O processamento de dados envolve diversas etapas para extrair insights relevantes. A coleta de posts através da API do Bluesky permite reunir um conjunto significativo de publicações para estudo. Em seguida, uma Word-Cloud é gerada para destacar visualmente as palavras mais frequentes, facilitando a identificação de termos recorrentes. A análise de correlação entre métricas de engajamento, como curtidas, comentários e compartilhamentos, auxilia na identificação de padrões de interação. A classificação de engajamento categoriza os posts com base em seu desempenho, evidenciando aqueles que obtiveram maior alcance e interação.

A previsão de engajamento é realizada utilizando o modelo ARIMA, que analisa o histórico de interações dos posts para estimar o desempenho futuro das publicações. A partir dos dados coletados, o modelo identifica tendências e padrões, permitindo prever o nível de engajamento esperado para novos posts. Além da previsão quantitativa, o sistema oferece sugestões sobre a postagem ideal, indicando o melhor dia da semana e horário para publicar, além da quantidade ideal de caracteres que tende a gerar maior interação. Essas recomendações visam otimizar o alcance e a eficácia das postagens, tornando a estratégia de conteúdo mais eficiente.

A análise de sentimentos utiliza o modelo VADER para classificar os posts como positivos, negativos ou neutros, exibindo uma seleção dos sentimentos mais extremos. Quando aplicável, um mapa interativo é gerado para visualizar a distribuição geográfica dos sentimentos, permitindo uma análise espacial do engajamento emocional. A modelagem de tópicos é realizada através do algoritmo Latent Dirichlet Allocation (LDA), que identifica os principais temas discutidos nos posts. Para melhor compreensão da evolução dos assuntos, a distribuição dos tópicos ao longo do tempo é visualizada, permitindo observar como os temas emergem e se transformam nas interações dos usuários.

A análise temporal do engajamento investiga como a interação dos usuários

varia ao longo do tempo, considerando fatores como o horário das postagens e o tamanho do texto. Essa exploração possibilita a identificação de padrões de engajamento, ajudando a compreender quais momentos e características das postagens influenciam a resposta do público.

4 Resultados

Utilizando o perfil do jornal americano New York Times para testar a aplicação, chegou aos resultados abaixo.

4.1 Análise Temporal do Engajamento

Os gráficos analisados mostram diferentes aspectos do engajamento:

- **Engajamento vs. Número de Caracteres:** O gráfico indica uma correlação positiva entre o número de caracteres e o engajamento. Postagens mais longas tendem a gerar maior interação.
- **Engajamento Médio por Hora do Dia:** Observa-se que o engajamento é maior entre 0h e 6h da manhã, caindo durante o dia e voltando a subir à noite.
- **Tendência do Engajamento ao Longo do Tempo:** Há um crescimento acentuado no engajamento ao longo dos meses, especialmente após dezembro.
- **Previsão do Engajamento Futuro:** A previsão sugere uma estabilização do engajamento em torno de 40.000 interações.

4.2 Análise de Sentimentos

O histograma da distribuição dos sentimentos usando o modelo VADER revela:

- A maioria dos posts tem um score de sentimento próximo de zero, indicando neutralidade.
- Há uma distribuição relativamente balanceada entre sentimentos positivos e negativos, com leve viés para sentimentos negativos.

4.3 Análise de Tópicos

A nuvem de palavras gerada para os tópicos principais destaca termos frequentemente associados às postagens:

- O nome **Trump** aparece em todos os tópicos, indicando que as postagens analisadas estão fortemente relacionadas a este assunto.
- Outros termos comuns incluem **president**, **government**, **federal** e **order**, sugerindo que o foco das discussões está em política e administração pública.

5 Conclusão

Este projeto fornece insights detalhados sobre os posts do NY Times e do Bluesky, contribuindo para a compreensão dos padrões de engajamento, a previsão do sucesso de postagens futuras e a identificação de temas predominantes. Além disso, ele pode ser expandido para outras redes sociais e refinado com modelos mais avançados de previsão e processamento de linguagem natural.