

RELATÓRIO DE ATIVIDADES 23 DE OUTUBRO

Projeto: Análise de Sentimentos no YouTube

Disciplina: Mineração de Texto

Data: 23/10/2025

Equipe:

Anderson de Matos Guimarães,

Gustavo Stefano Thomazinho,

Renan Ost

1. OBJETIVO

Consolidar as Fases 1 (Coleta) e 2 (Análise) com melhorias de robustez, desempenho e compatibilidade com Flask. Os notebooks finais são parametrizados, sem inputs interativos, e geram arquivos com timestamp para preservação do histórico.

2. DECISÕES TÉCNICAS

Deteção automática do tipo de entrada (canal/vídeo) a partir do padrão da URL/ID.

Parâmetros universais: MAX_VIDEOS (apenas para canal) e MAX_COMMENTS (por vídeo) — padrão 5 e 1000.

Coleta com paginação via nextPageToken; comentários top-level.

Exportação com timestamp:

comentarios_coletados_YYYYMMDD_HHMMSS.csv e

comentarios_analisados_YYYYMMDD_HHMMSS.csv.

Limpeza textual robusta (regex + emoji.demojize) e filtro de idioma com langdetect.

Modelo BERT-PT (HuggingFace) com inferência em batch (batch_size=32) e truncation habilitado.

Padronização de rótulos de sentimento para: positivo, negativo, neutro.

EDA básica para sanidade (histograma de tamanho; distribuição e comparativo por vídeo).

Compatibilidade total com Flask (sem input()); funções para serem chamadas por rotas).

3. TESTES REALIZADOS

Executamos cenários de coleta para canal (top 5 vídeos) e vídeo único com limites ajustáveis de comentários. Verificamos a integridade dos CSVs, a eliminação de duplicidades e a coerência dos rótulos de sentimento. A inferência em lote reduziu significativamente o tempo de processamento em bases médias.

4. ESTRATÉGIA: PRÓXIMA FASE (FASE 3 — VISUALIZAÇÃO E EDA)

Criar o notebook `visualizacao.ipynb` para consolidar dashboards interativos (Plotly) e análises textuais (WordCloud, frequências), consumindo os arquivos `comentarios_analisados_*.csv`. Definir métricas e painéis comparativos por vídeo/canal e preparar a transição para Flask (Fase 4).

5. CONCLUSÃO

As Fases 1 e 2 estão consolidadas conforme boas práticas de mineração de textos. O pipeline está modular, reproduzível e pronto para integração na aplicação web.