**Slide 1 – Capa**

**Conteúdo:** *Análise de Sentimentos em Comentários do YouTube* + nomes do grupo.  
**Fala (Anderson):**

“Boa tarde. Nosso projeto tem como tema a análise de sentimentos em comentários do YouTube. Somos o grupo Anderson, Renan e Gustavo, e hoje vamos apresentar os avanços da primeira etapa, focando nos critérios definidos pelo professor.”

**Slide 2 – Relevância do Projeto**

**Conteúdo:** importância acadêmica, prática e aplicabilidade.  
**Fala (Anderson):**

“Escolhemos esse tema porque os comentários do YouTube são uma rica fonte de informações. A análise desses dados permite compreender as reações do público, algo relevante para empresas, criadores de conteúdo e pesquisas acadêmicas. Assim, o trabalho une valor prático e científico.”

**Slide 3 – Estrutura do Projeto**

**Conteúdo:** fluxo das etapas: Coleta → Pré-processamento → Metodologia → Resultados → Discussão → Conclusão.  
**Fala (Anderson):**

“A estrutura do nosso trabalho segue essa linha lógica: iniciamos com a coleta, tratamos os dados, aplicamos a metodologia, analisamos os resultados, discutimos as limitações e finalizamos com as considerações. Essa organização nos garante clareza e continuidade.”

**Slide 4 – Dados Minerados**

**Conteúdo:** origem (YouTube API), tipo de dados e amostra inicial.  
**Fala (Renan):**

“Utilizamos a API do YouTube para coletar comentários de vídeos com alto engajamento. Essa base inicial é o ponto de partida para nossa análise. Aqui temos informações brutas como autor, texto, data e vídeo de origem, que servirão para o pré-processamento.”

**Slide 5 – Pré-processamento**

**Conteúdo:** limpeza, normalização, stopwords, tokenização, exemplo antes/depois.  
**Fala (Renan):**

“Após a coleta, é necessário preparar os dados. Realizamos etapas de limpeza, como remover links, emojis e caracteres especiais, além de normalizar o texto em minúsculas. Também aplicamos remoção de stopwords e tokenização. Este exemplo mostra como um comentário bruto é transformado em uma versão mais limpa e útil para o modelo.”

**Slide 6 – Metodologia**

**Conteúdo:** TF-IDF, modelos clássicos (Naive Bayes e Regressão Logística), futuro com embeddings.  
**Fala (Renan):**

“Para transformar os textos em representações numéricas, utilizamos o TF-IDF. Como classificadores, escolhemos inicialmente Naive Bayes e Regressão Logística por serem rápidos e eficientes. Mais adiante, pretendemos aplicar métodos avançados como embeddings, incluindo Word2Vec e BERT, para capturar contexto semântico.”

**Slide 7 – Resultados Parciais**

**Conteúdo:** relatório inicial de sentimentos (ainda simples).  
**Fala (Gustavo):**

“Mesmo em fase inicial, já conseguimos gerar relatórios parciais dos sentimentos. Esses resultados mostram que o pipeline está funcionando, ainda que em pequena escala. Eles servem como prova de conceito e indicam que estamos no caminho certo.”

**Slide 8 – Resultados Esperados**

**Conteúdo:** métricas (acurácia, precisão, recall, F1-score), distribuição de sentimentos prevista.  
**Fala (Gustavo):**

“Os resultados esperados incluem métricas de desempenho, como acurácia, precisão, recall e F1-score. Também projetamos gráficos que mostrarão a distribuição de sentimentos em positivos, negativos e neutros. Isso permitirá avaliar a qualidade do modelo e a utilidade prática da análise.”

**Slide 9 – Discussão Crítica e Limitações**

**Conteúdo:** limitações (dataset pequeno, classes desbalanceadas, modelos simples) e próximos passos.  
**Fala (Gustavo):**

“Reconhecemos que a base ainda é pequena e pode estar desbalanceada. Além disso, os modelos escolhidos são básicos e podem ter limitações. Por isso, nossos próximos passos incluem ampliar o dataset, aplicar técnicas de balanceamento e explorar modelos mais robustos, como embeddings avançados. Isso garantirá resultados mais consistentes na próxima etapa.”