



ANÁLISE DE SENTIMENTOS

PROCESSAMENTO DE LINGUAGEM NATURAL

DAYLLAN, ERIC, GABRIELA, GIOVANNA, LUCAS E MICHEL

Agenda

- **ANDAMENTO DO PROJETO**
Sprint's anteriores
- **OBJETIVO DA SPRINT**
Artefatos desenvolvidos
- **PRÉ-PROCESSAMENTO**
Como os dados foram tratados?
- **ANÁLISE PRELIMINAR**
Primeiros insights dos dados
- **CRONOGRAMA**
Passos Futuros

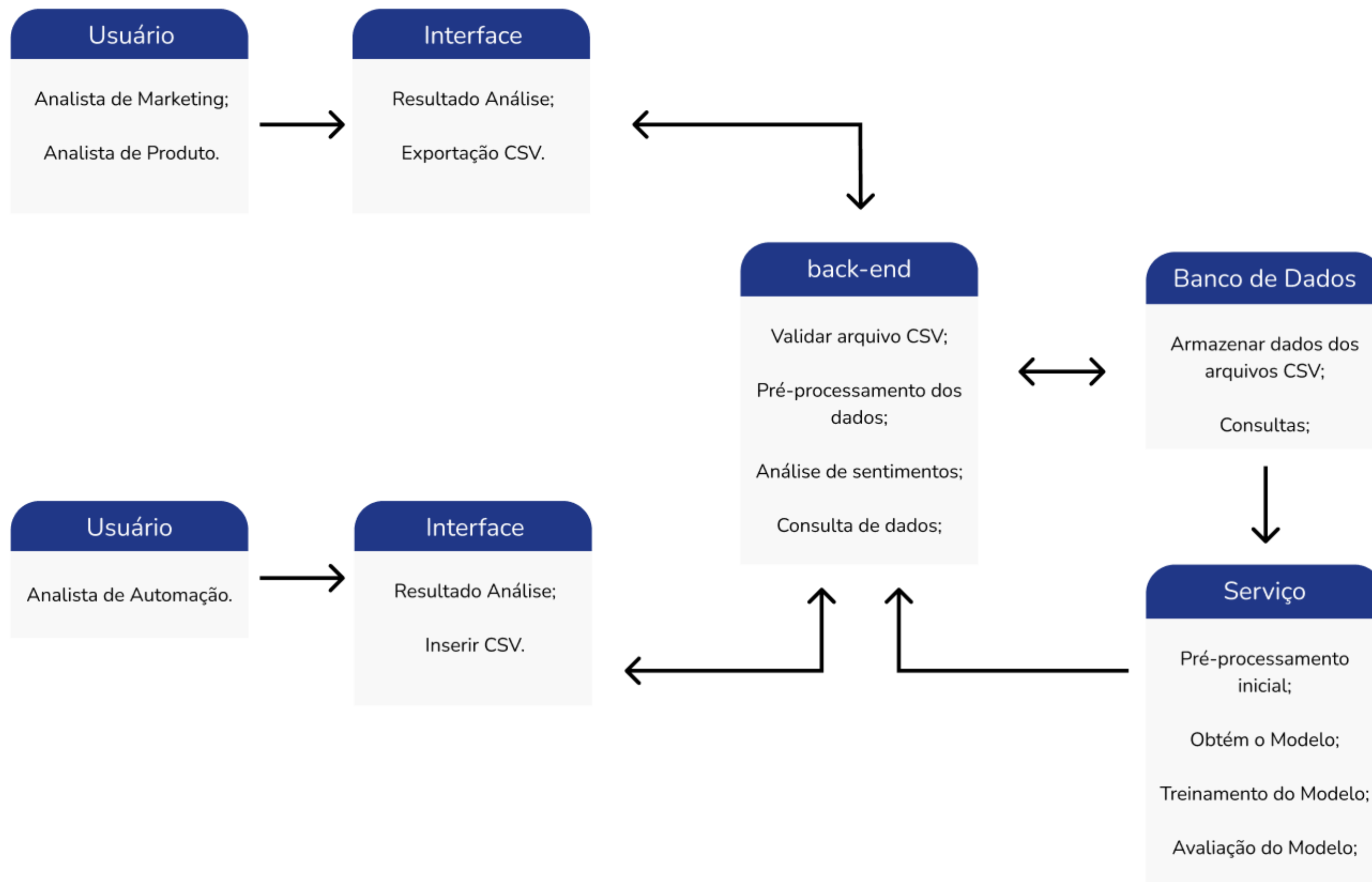


Andamento do Projeto

Arquitetura
Proposta



Arquitetura Macro da solução



Itens Entregáveis

Objetivo da Sprint 2 : Pré-processamento e Bag Of Words

- 1 **Pré-Processamento dos dados** : Realizar a organização, limpeza e ajustes da base de dados disponibilizada.
- 2 **Modelo de Bag of Words**: Iniciar a programação e teste com o modelo.



Pré-processamento dos dados

- 1 **Remoção de comentário "BTG Pactual":** Nesta etapa do processo estamos focando em analisar os usuários.
- 2 **Anonimização:** Na coluna "autor" onde se encontrava o @ do usuário, foi colocado um índice numérico.
- 2 **Formatação Data:** A data foi formatada em dd-mm-yyyy, para filtragem dos comentários em até três meses.



Pré-processamento dos dados



- 1 **Tokenização:** Individualização das palavras para a análise posterior.
- 2 **Remoção de ruído:** Tratamento de emojis, remoção de caracteres especiais, conversão para minúsculas e pontuações.
- 3 **StopWords:** Remoção das palavras de ligação que não seriam úteis na análise de sentimento.
- 4 **Vetorização:** A transformar uma entrada de dados, em um vetor de números.

Modelo Bag Of Words

BoW é uma forma de representar o texto de acordo com a ocorrência das palavras nele. Não leva em conta a ordem ou a estrutura das palavras no texto, apenas se ela aparece ou a frequência com que aparece nele. Passos para aplicação:

- 1) Selecionar os dados;
- 2) Gerar o vocabulário;
- 3) Formar vetores a partir do documento;



Pipeline

2. Normalização e Remoção de ruídos

banco incrível, melhor escolha
que fiz **:rosto_apaixonado:**

4. Remoção StopWords

"banco", "incrível", "melhor",
"escolha", "fiz",
":rosto_apaixonado:"

1. Dados (Input)

Banco incrível, melhor
escolha que fiz 😊

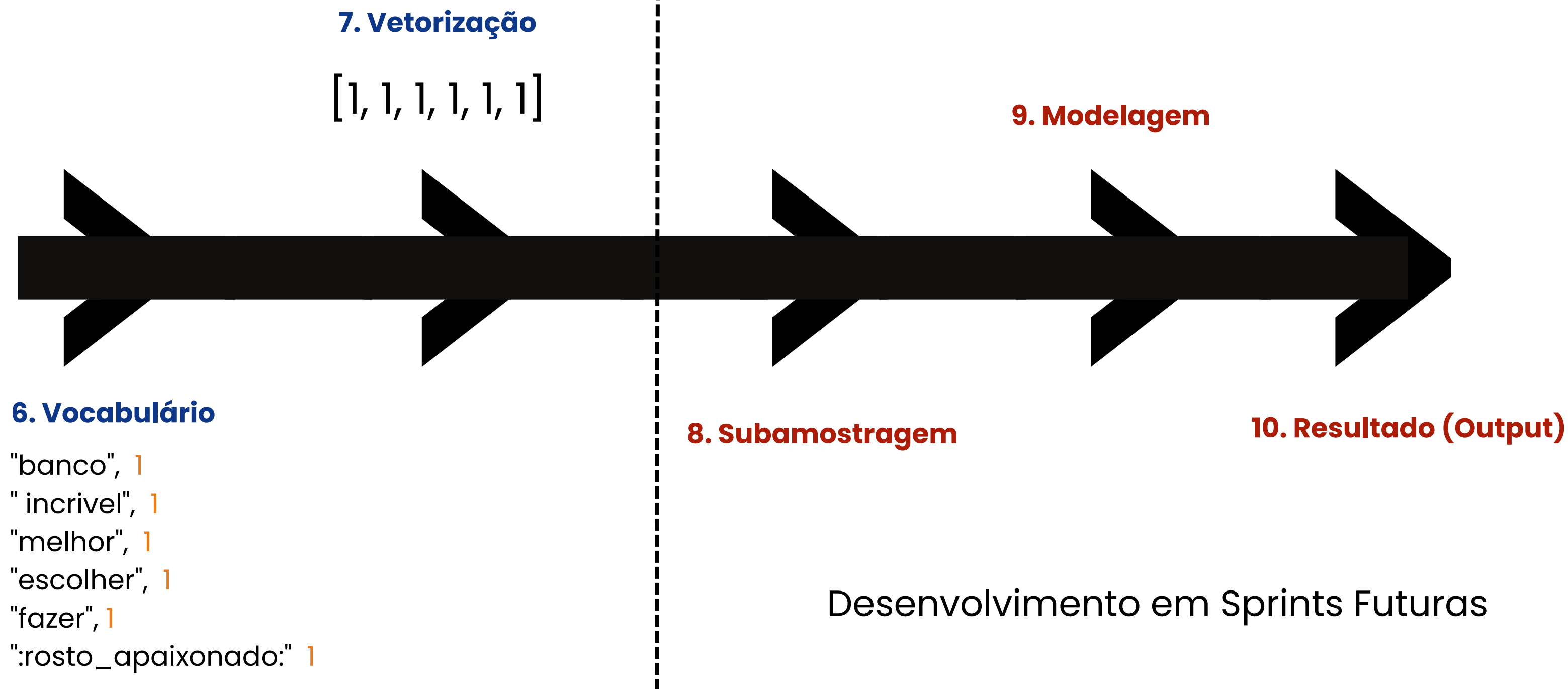
3. Tokenização

"banco", "incrível", "melhor",
"escolha", "que", "fiz",
":rosto_apaixonado:"

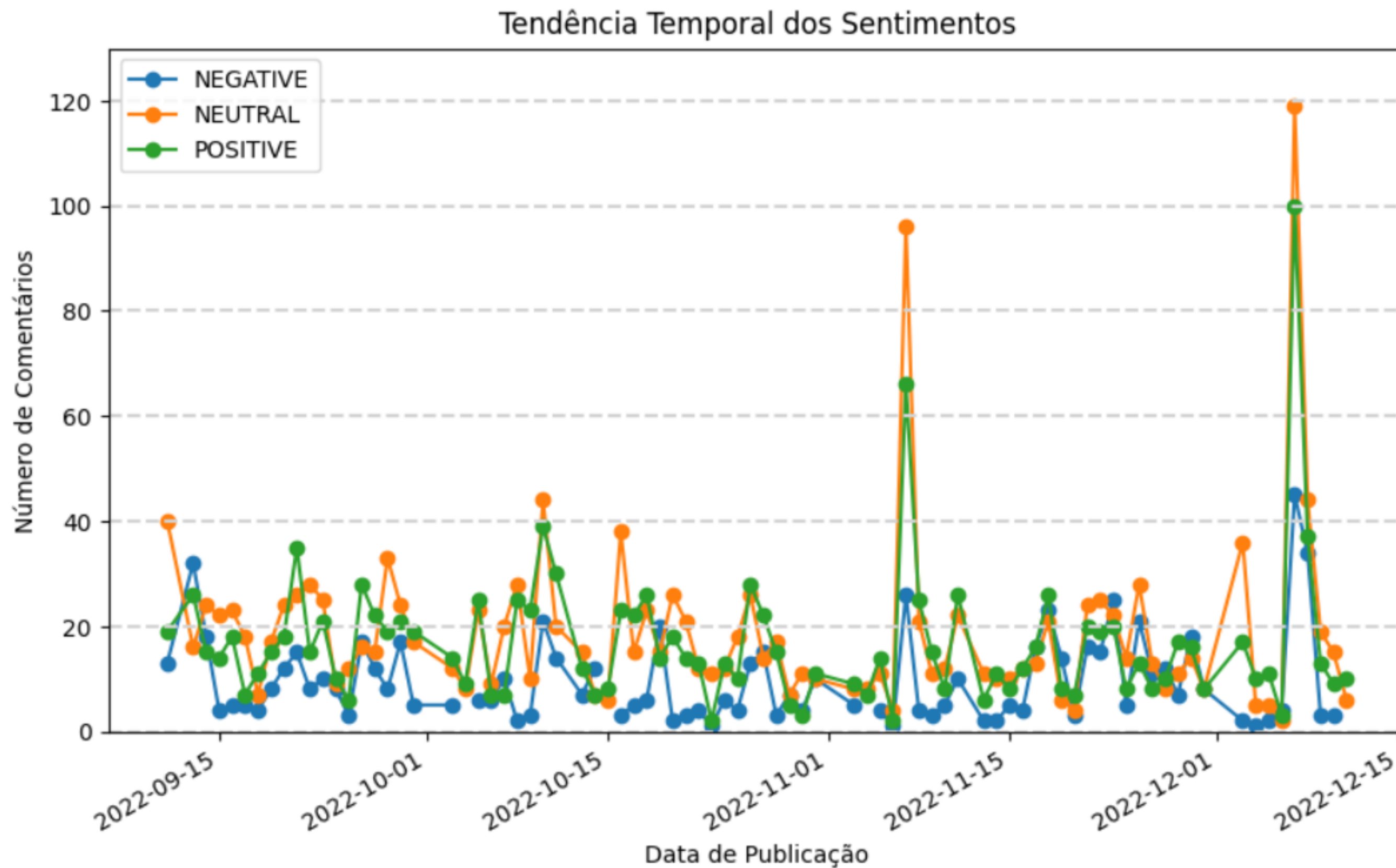
5. Verbos para infinitivo

"banco", "incrível", "melhor",
"escolher", **fazer**,
":rosto_apaixonado:"

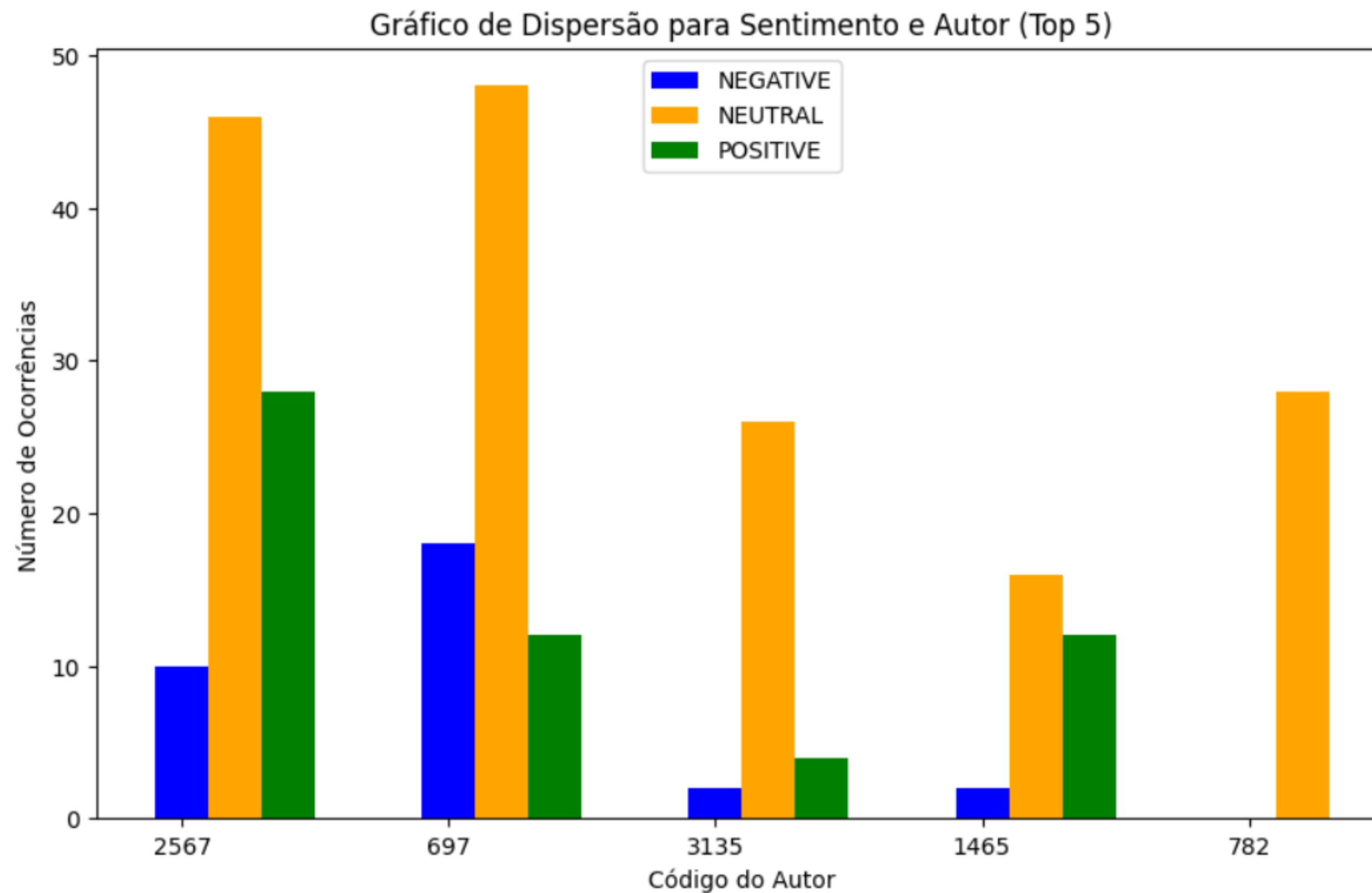
Pipeline



Análise Preliminar



Análise Preliminar





[illegible]

Próximos Passos

Conversão de palavras em um vetor
de baixa dimensão.

Modelo Word2Vec

Naive Bayes

Tabela de probabilidades a partir da
técnica de classificação de dados.

Cronograma Projeto

CRONOGRAMA		
1	Recebimento da base de dados;	CONCLUÍDO
2	Validação de Premissas;	CONCLUÍDO
3	Entendimento dos dados;	CONCLUÍDO
4	Pré-processamento dos dados;	CONCLUÍDO
5	Modelagem - Modelo de Bag of Words;	CONCLUÍDO
6	Modelo Word2Vec Análise de sentimentos;	PENDENTE
7	Prototipagem da interface;	PENDENTE
8	Criação do back-end;	PENDENTE
9	Deploy do melhor modelo;	PENDENTE
10	Entrega Final;	PENDENTE

TIME

+



Dayllan
de Souza Alho

+



Eric
Tachdjian

+



Gabriela
de Moraes Silva

+



Giovanna
Furlan Torres

+



Lucas
de Britto

+



Michel
Mansur

Obrigado(a)!

Estamos disponíveis para maiores esclarecimentos.

