



ANÁLISE DE SENTIMENTOS

PROCESSAMENTO DE LINGUAGEM NATURAL

DAYLLAN, ERIC, GABRIELA, GIOVANNA, LUCAS E MICHEL

Agenda

- **OBJETIVO DA SPRINT**
Artefatos desenvolvidos
- **ARQUITETURA MACRO**
Solução
- **EXECUÇÃO DOS ARTEFATOS**
Resultados
- **CRONOGRAMA**
Passos Futuros



Andamento do Projeto



Itens Entregáveis

Objetivo da Sprint 3 : Word2Vec e Modelos de Treinamento

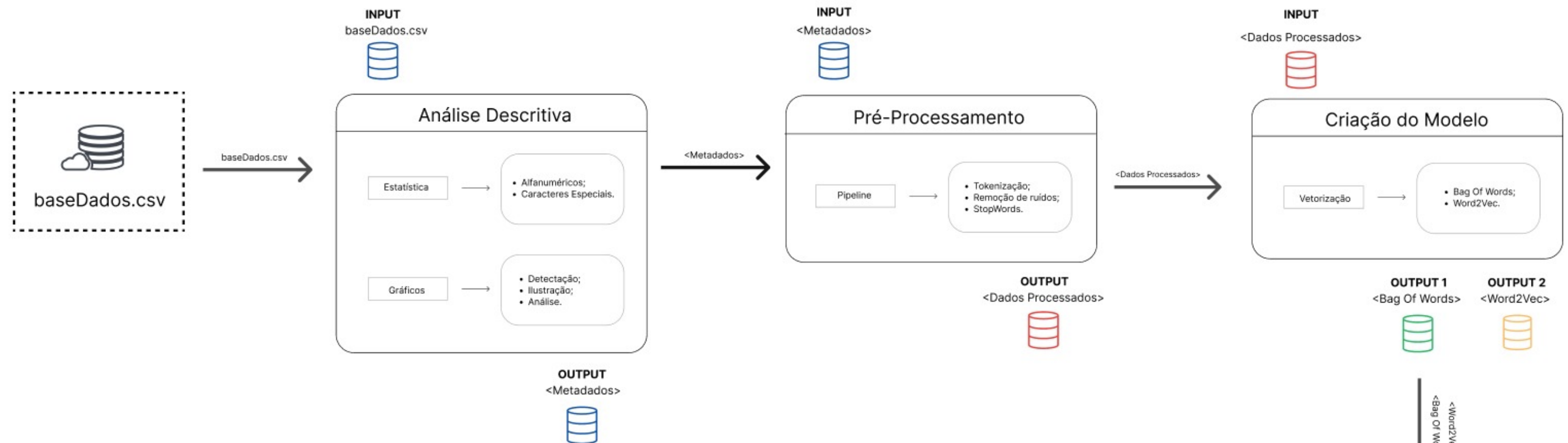
1

Word2Vec : modelo de aprendizado de máquina que representa palavras como vetores.

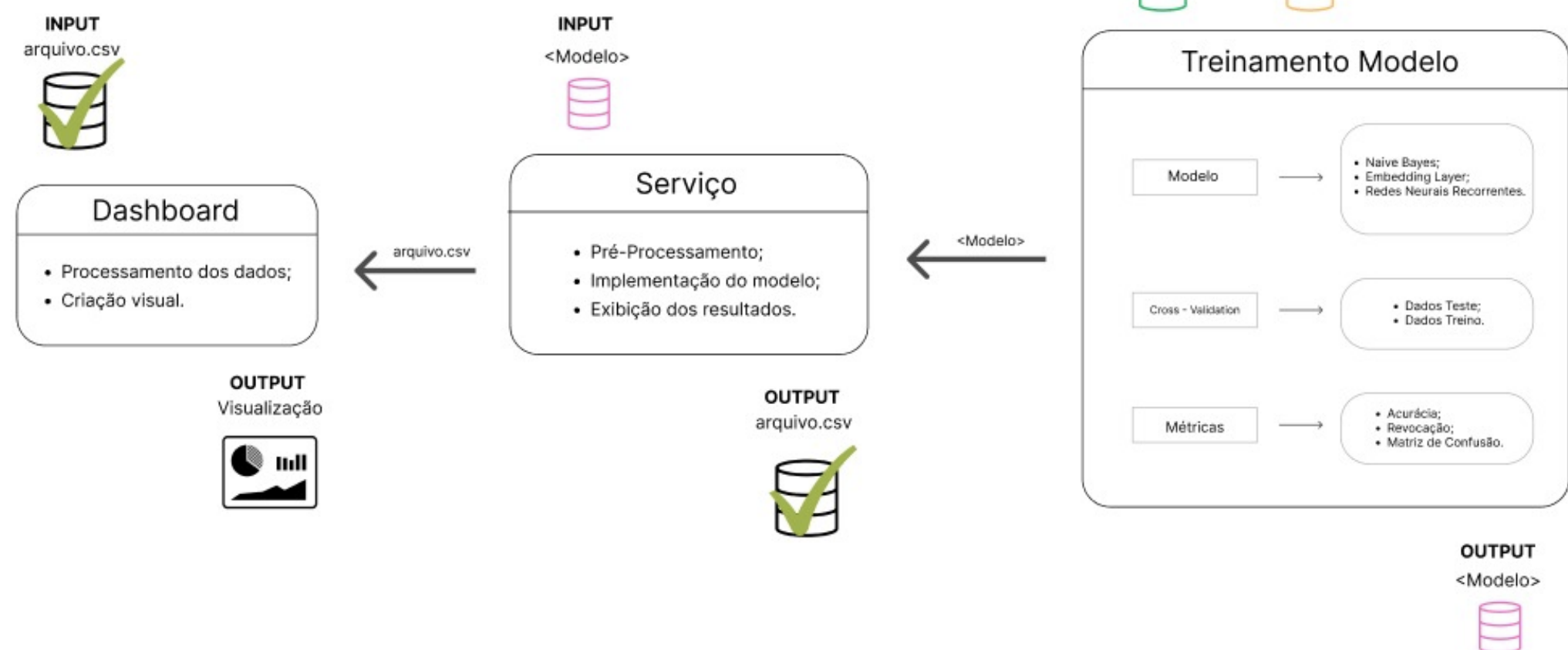
2

Modelos de Treinamento : Naive Bayes Gaussiana, RNN e Embedding Layer.

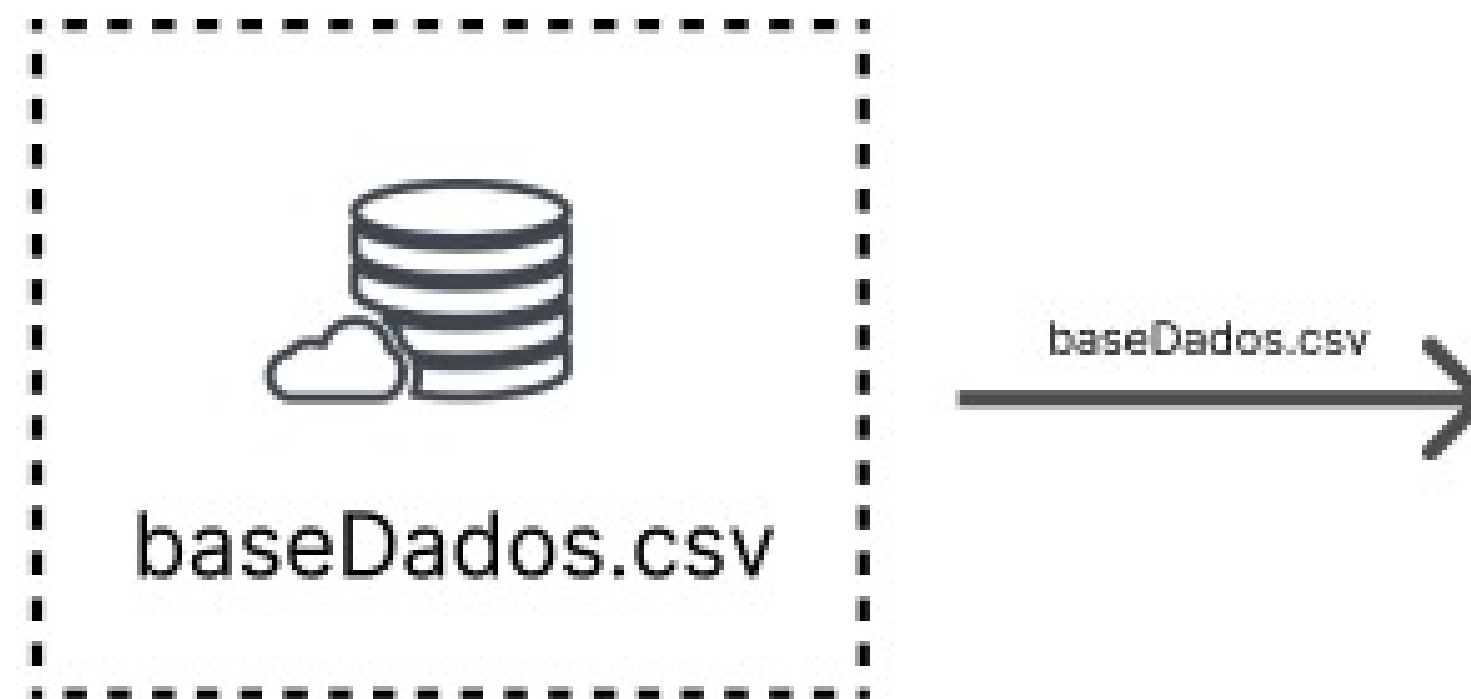




Arquitetura Macro da solução

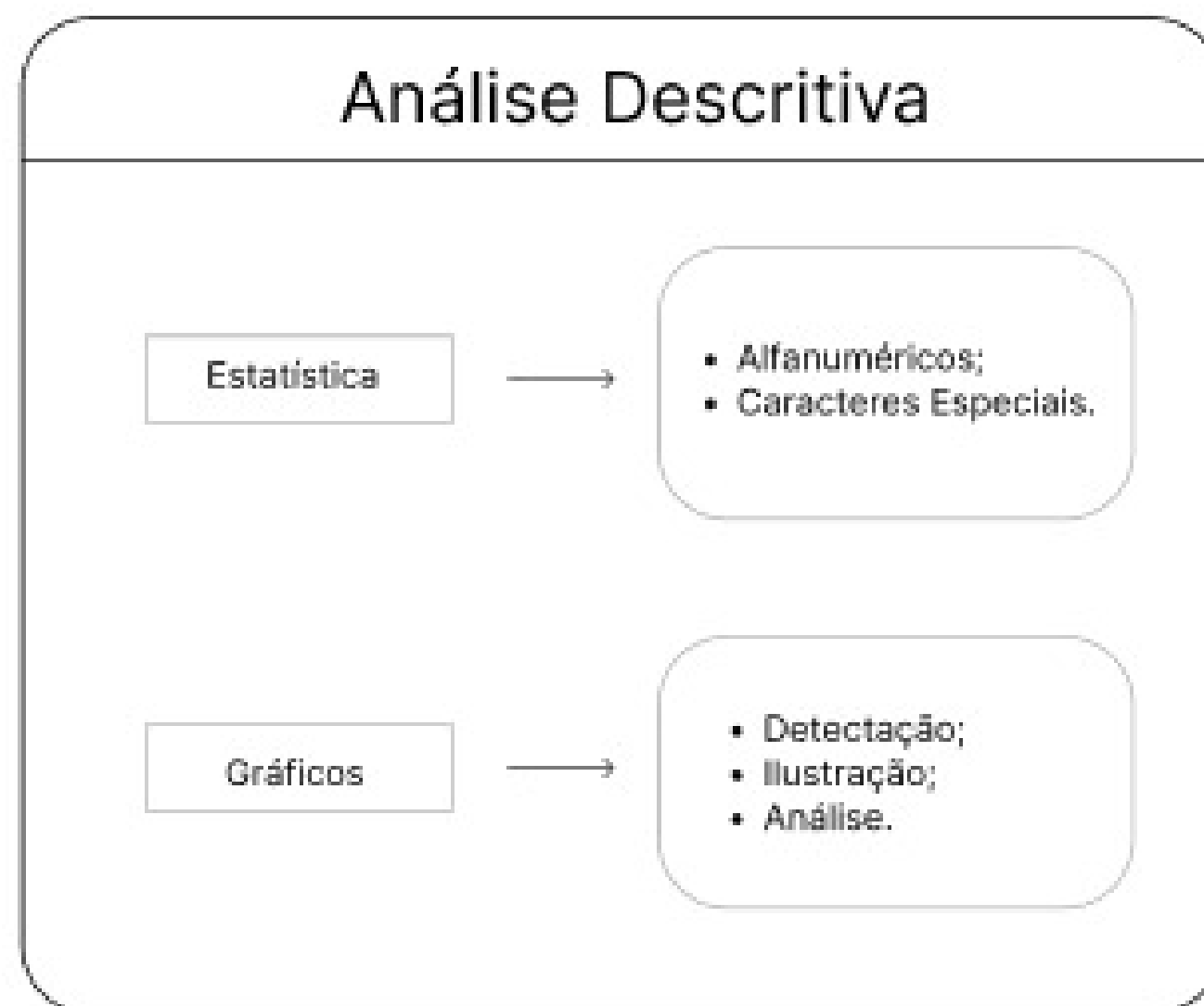


Origem dos Dados



Análise Descritiva

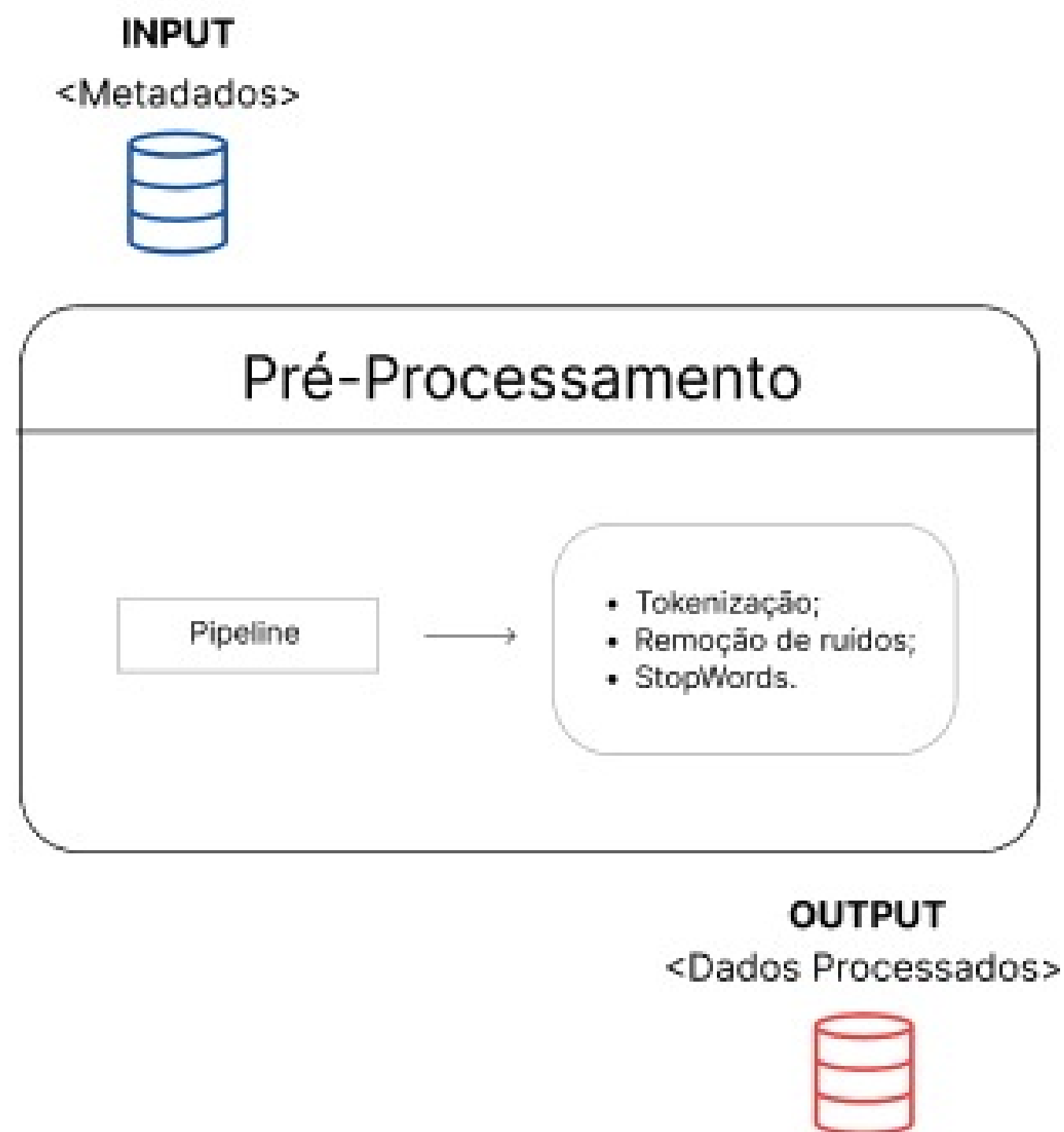
INPUT
baseDados.csv



OUTPUT
<Metadados>



Pré-Processamento



Criação do Modelo

INPUT

<Dados Processados>



Criação do Modelo

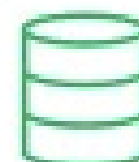
Vetorização



- Bag Of Words;
- Word2Vec.

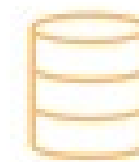
OUTPUT 1

<Bag Of Words>



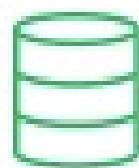
OUTPUT 2

<Word2Vec>

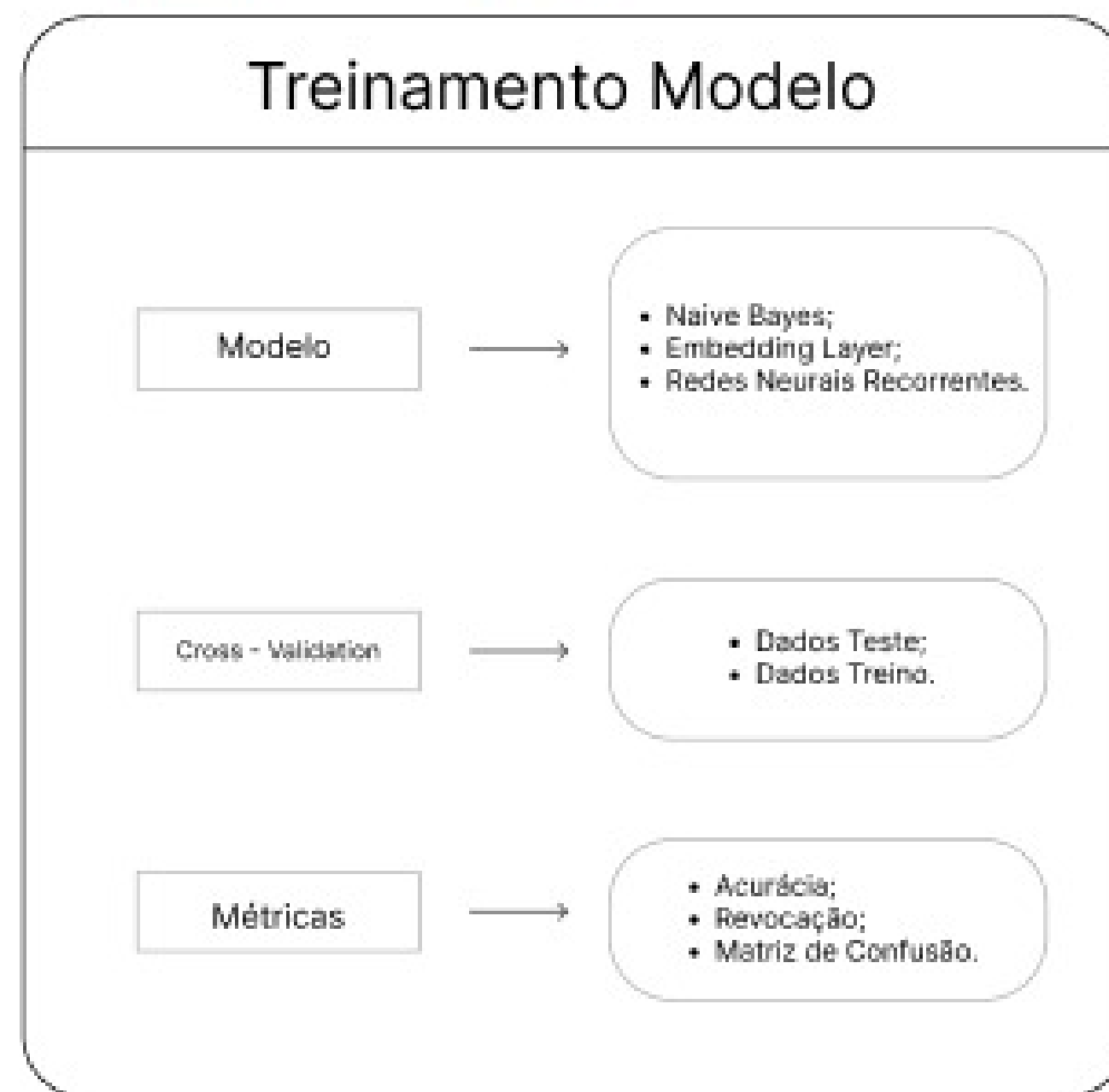
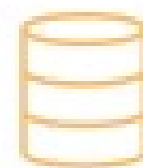


Treinamento do Modelo

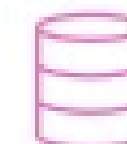
INPUT 1
<Bag Of Words>



INPUT 2
<Word2Vec>

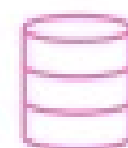


OUTPUT
<Modelo>



Serviço

INPUT
<Modelo>



Serviço

- Pré-Processamento;
- Implementação do modelo;
- Exibição dos resultados.

OUTPUT
arquivo.csv



Dashboard

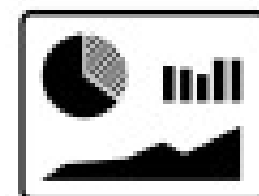
INPUT
arquivo.csv



Dashboard

- Processamento dos dados;
- Criação visual.

OUTPUT
Visualização



Modelo Word2Vec

Modelo que **transforma palavras em vetores**, **capturando** suas características **semânticas**, definições e **contexto**. O modelo **agrupa vetores de palavras similares** e fornece estimativas fortes sobre seus significados com **base em sua ocorrência nos textos**.

Passos para aplicação:

- 1) Construção do modelo Word2Vec;
- 2) Gerar vetores de palavras;
- 3) Exploração e análise dos vetores;



Bag of Words **V****S** Word2Vec

Representação baseada em contagem: considera **apenas a frequência** das palavras, **sem considerar contexto**, palavras tratadas de forma **independente**.

Alta dimensionalidade: gera vetores de alta dimensionalidade, onde **cada dimensão representa uma palavra única** no vocabulário.

Perda de informações semânticas e contextuais: **palavras diferentes** com **significados semelhantes** podem ter **vetores** completamente **diferentes**.

Representação densa e vetores de palavras: as **palavras com significados semelhantes** têm **vetores próximos** no espaço de vetores. Essa representação captura melhor as **relações semânticas** e permite operações matemáticas entre os vetores de palavras.

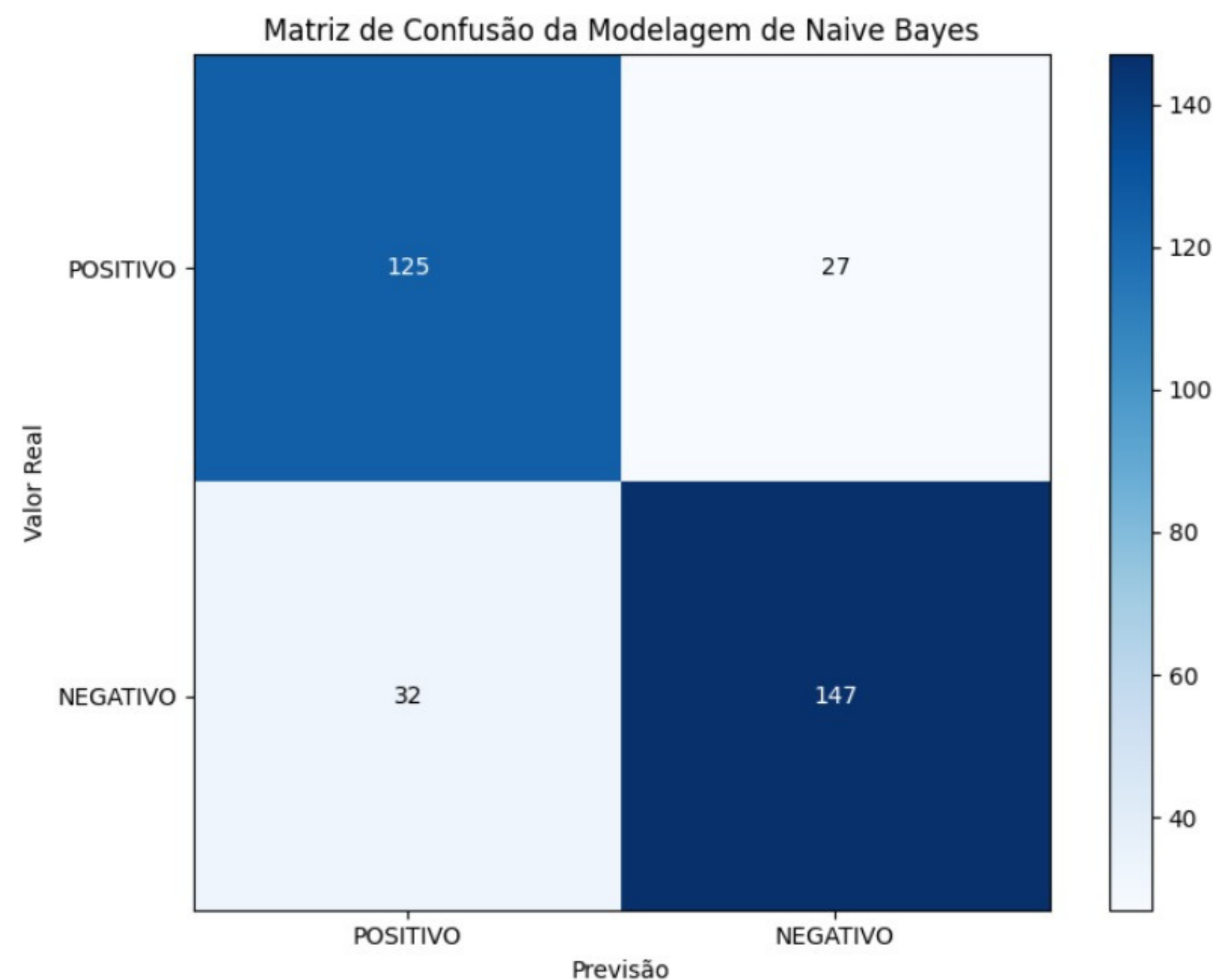
Redução da dimensionalidade: Os **vetores** são geralmente de **tamanho fixo**, independentemente do tamanho do vocabulário, o que torna mais eficiente em termos computacionais e lida melhor com problemas de alta dimensionalidade..

Modelo Naive Bayes Gaussiana (word2vec)

Classifica dados numéricos contínuos; calcula a probabilidade de pertencer a cada classe. É simples de implementar e pode funcionar bem em muitos casos, embora faça a simplificação de assumir **independência entre os atributos**.

Resultados:

Acurácia: 0,81

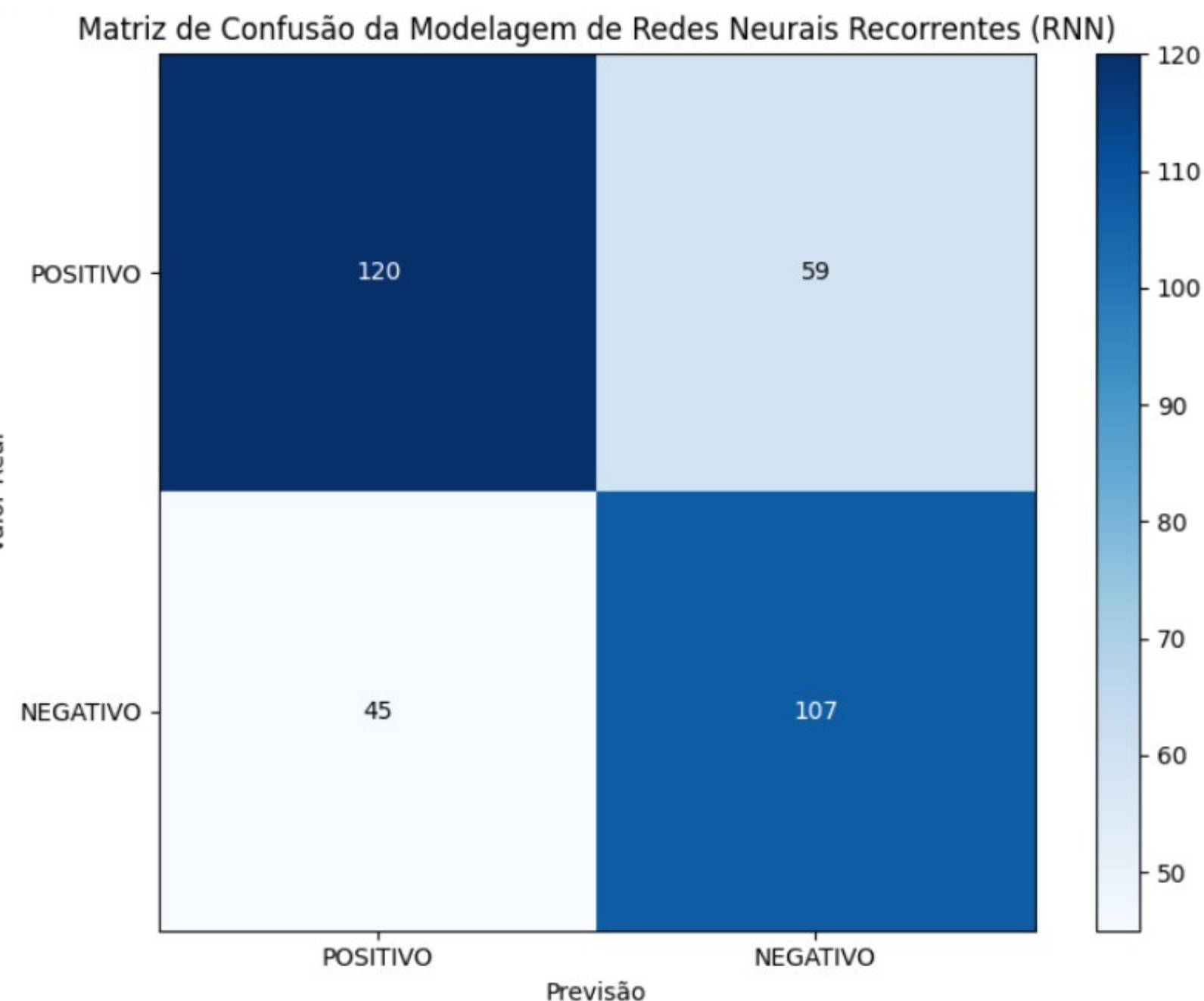


Modelo RNN (Word2Vec)

Modelo de aprendizado de máquina usado para **processar dados sequenciais**, como texto ou séries temporais. Ele possui **conexões em loop**, permitindo que **informações anteriores influenciem as próximas etapas**. Isso o torna adequado para capturar dependências de longo prazo em sequências.

Resultados:

Acurácia: 0,75

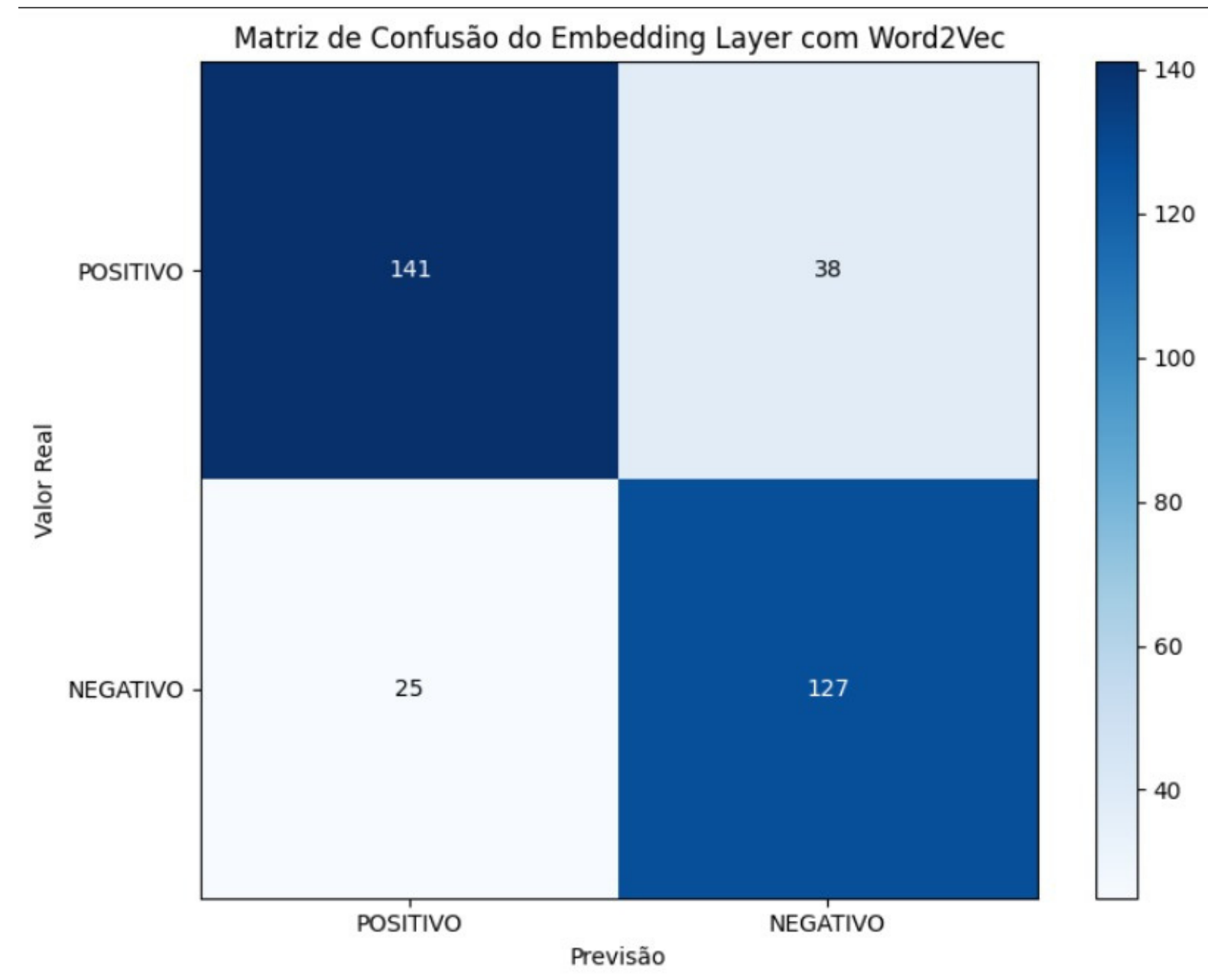


Modelo Embedding Layer (Word2Vec)

Parte de uma **rede neural** que mapeia palavras ou itens em um espaço vetorial de dimensão reduzida. Ela representa **cada palavra** ou item por meio de **um vetor** denso. Essa representação permite que o modelo **capture relações semânticas** e contextuais entre as palavras.

Resultados:

Acurácia: 0,81



Próximos Passos

Testar com novos hiperparâmetros
e features.

**Aprimorar
resultados**

**Investigação de
novos Modelos**

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; | EM ANDAMENTO |
| 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.

