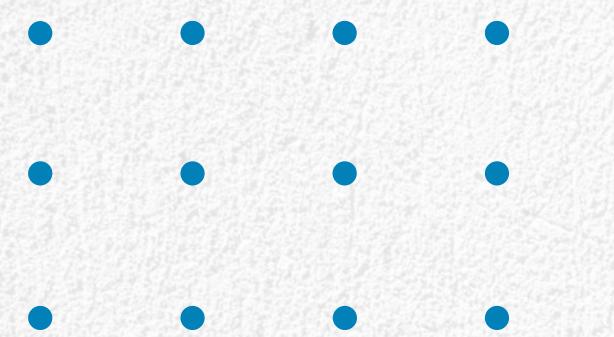






Jäger

Sprint 3 - Desenvolver modelos Word2Vec
com Naive Bayes e Bedding Layer



Revisão da Sprint 2

Bag of Words (BoW)

- Técnica que transforma documentos em vetores.
- Vetorizador de Contagem: **leva em conta frequência** de uma palavra no texto.
- Ajusta a frequência das palavras levando em conta a **importância delas no documento** e no corpus inteiro.

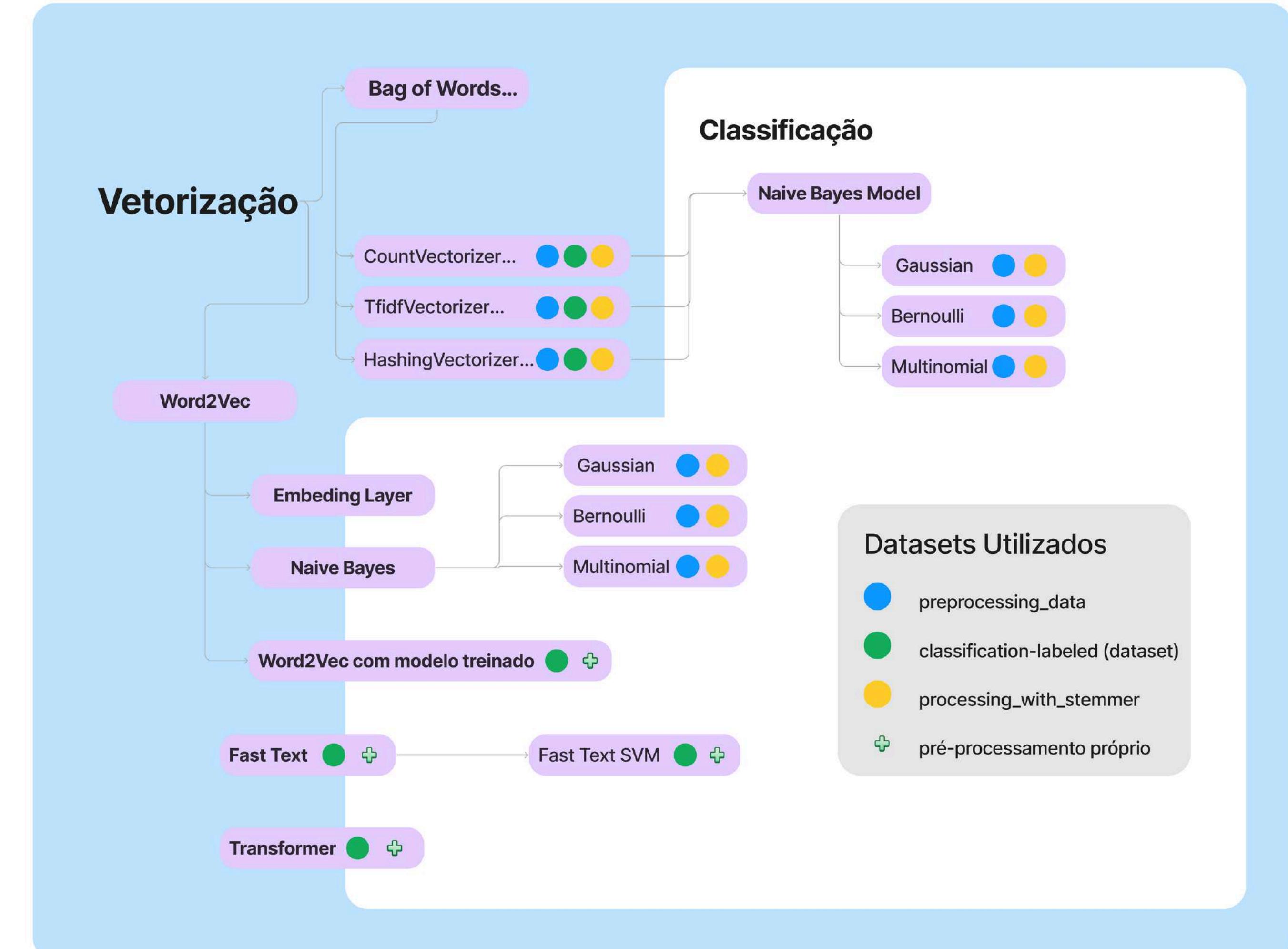
Naive Bayes

- Classificador probabilístico baseado no Teorema de Bayes
- Assume independência entre as características.
- **Naive Bayes Multinomial e Normal:** demonstraram um desempenho sólido em nosso conjunto de dados.

Métricas

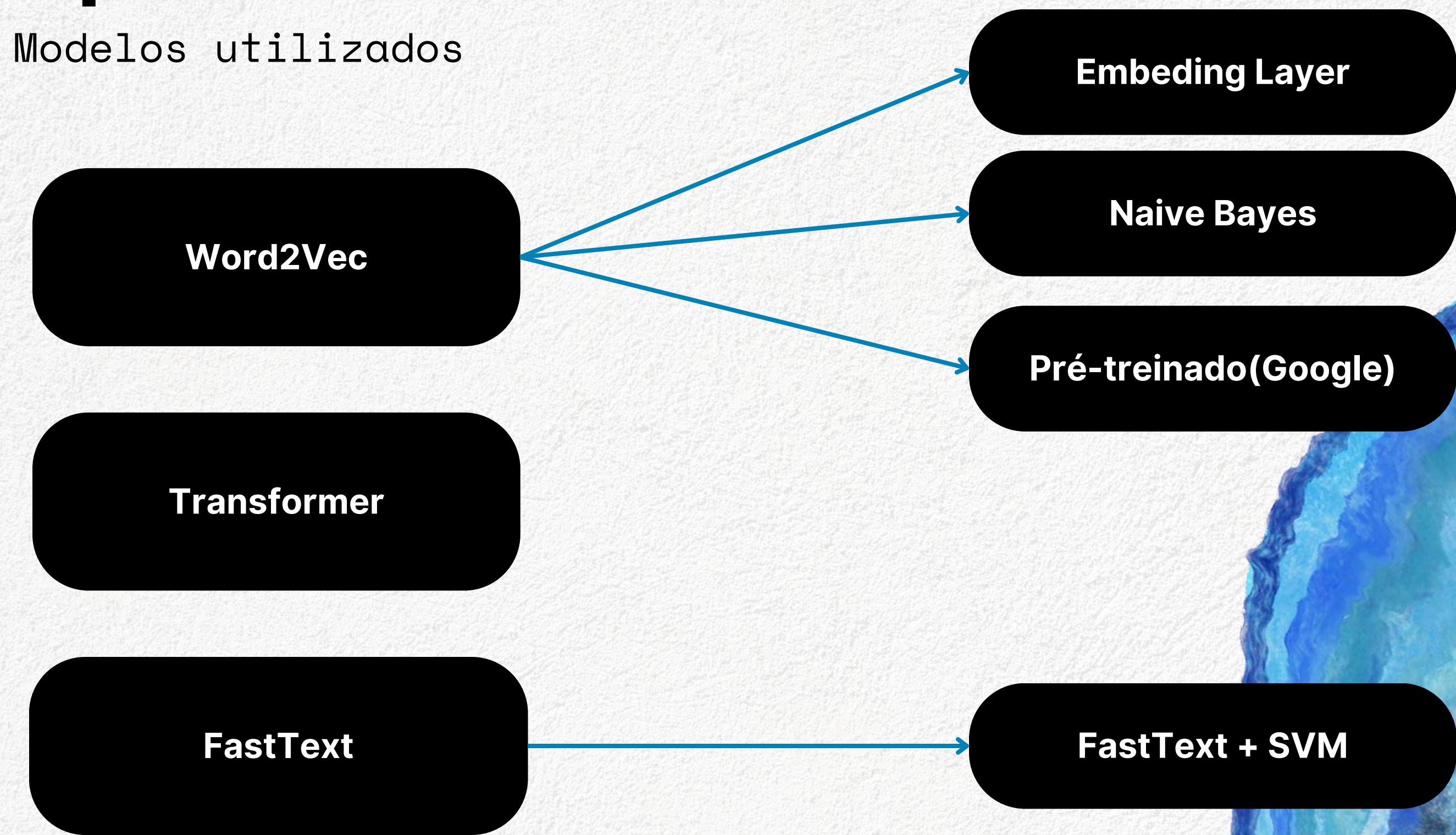
No sprint anterior, o modelo Bag of Words alcançou uma **precisão de 64,82%**, enquanto o modelo **Naive Bayes atingiu 64,72%**. Esses resultados forneceram uma base sólida para nossos esforços de classificação de texto.

Visão geral dos modelos



Sprint 3

Modelos utilizados



Boot

Sprint 3

Modelos utilizados

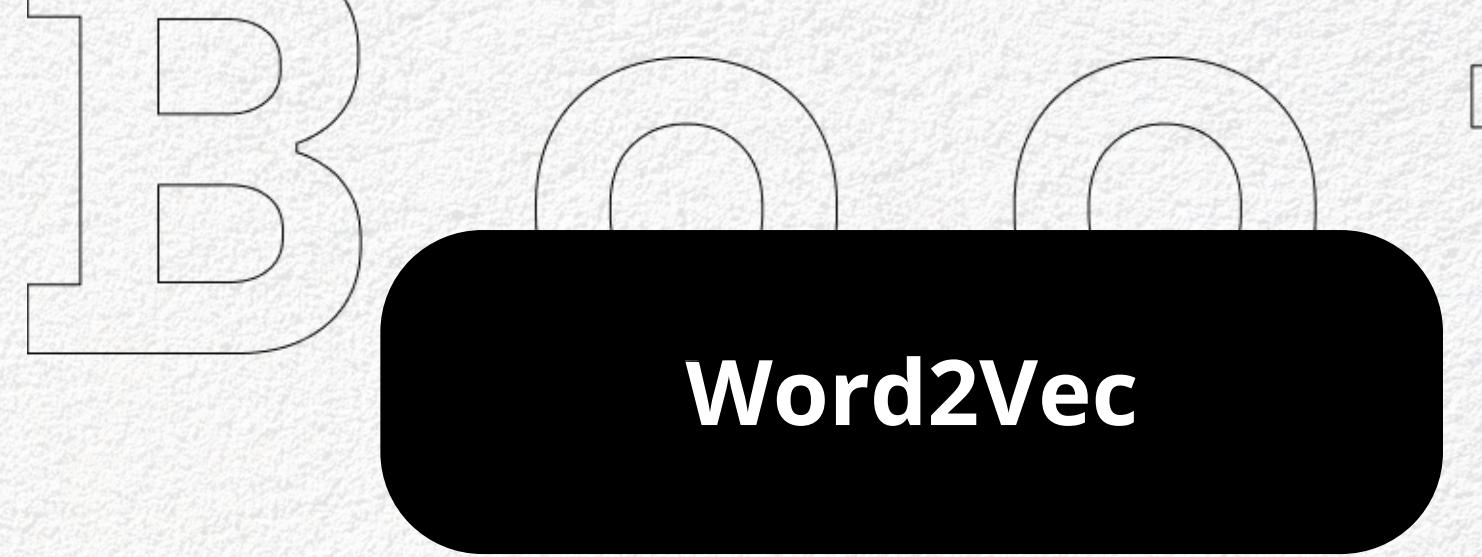
Word2Vec

Word2Vec é uma técnica que transforma palavras em vetores contínuos, capturando informações semânticas e contextuais.

Boot

Sprint 3

Modelos utilizados



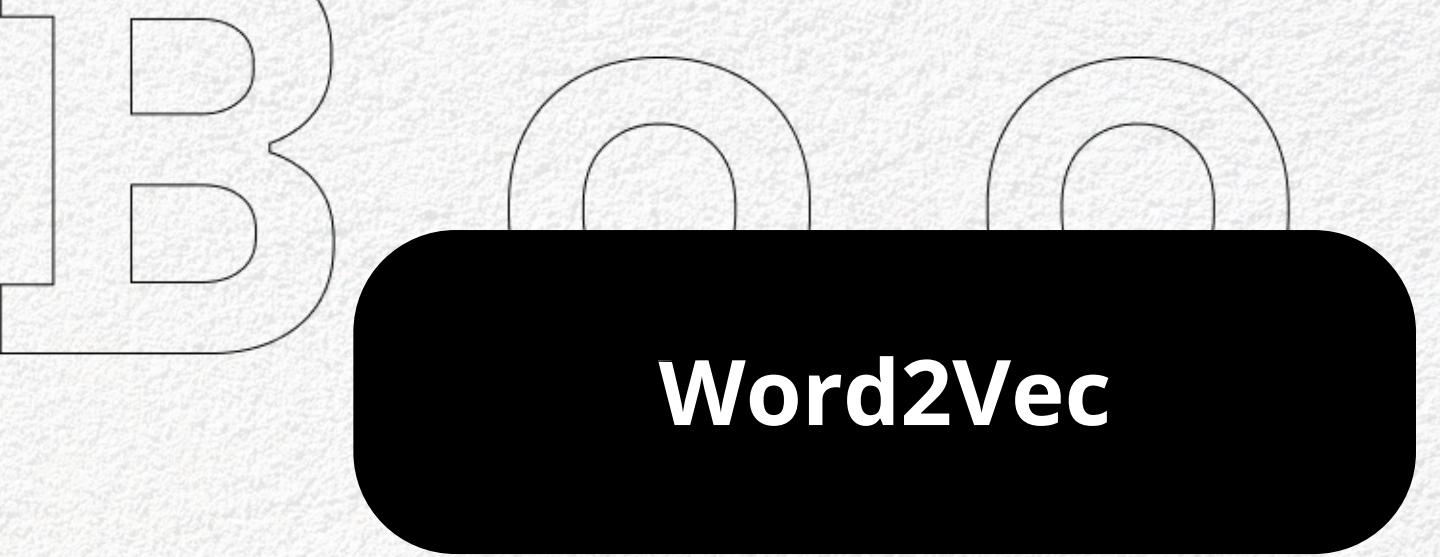
Word2Vec + Embedding Layer

Word2Vec + Naive Bayes

Boot

Sprint 3

Modelos utilizados



Word2Vec + Embedding Layer

A camada de embedding em redes neurais, como usada no Word2Vec, é uma camada de rede neural que transforma palavras em vetores densos de uma dimensão fixa. Esse processo de mapeamento é crucial para [converter dados textuais](#) em uma [forma numérica](#) que as redes neurais podem processar.

Resultados

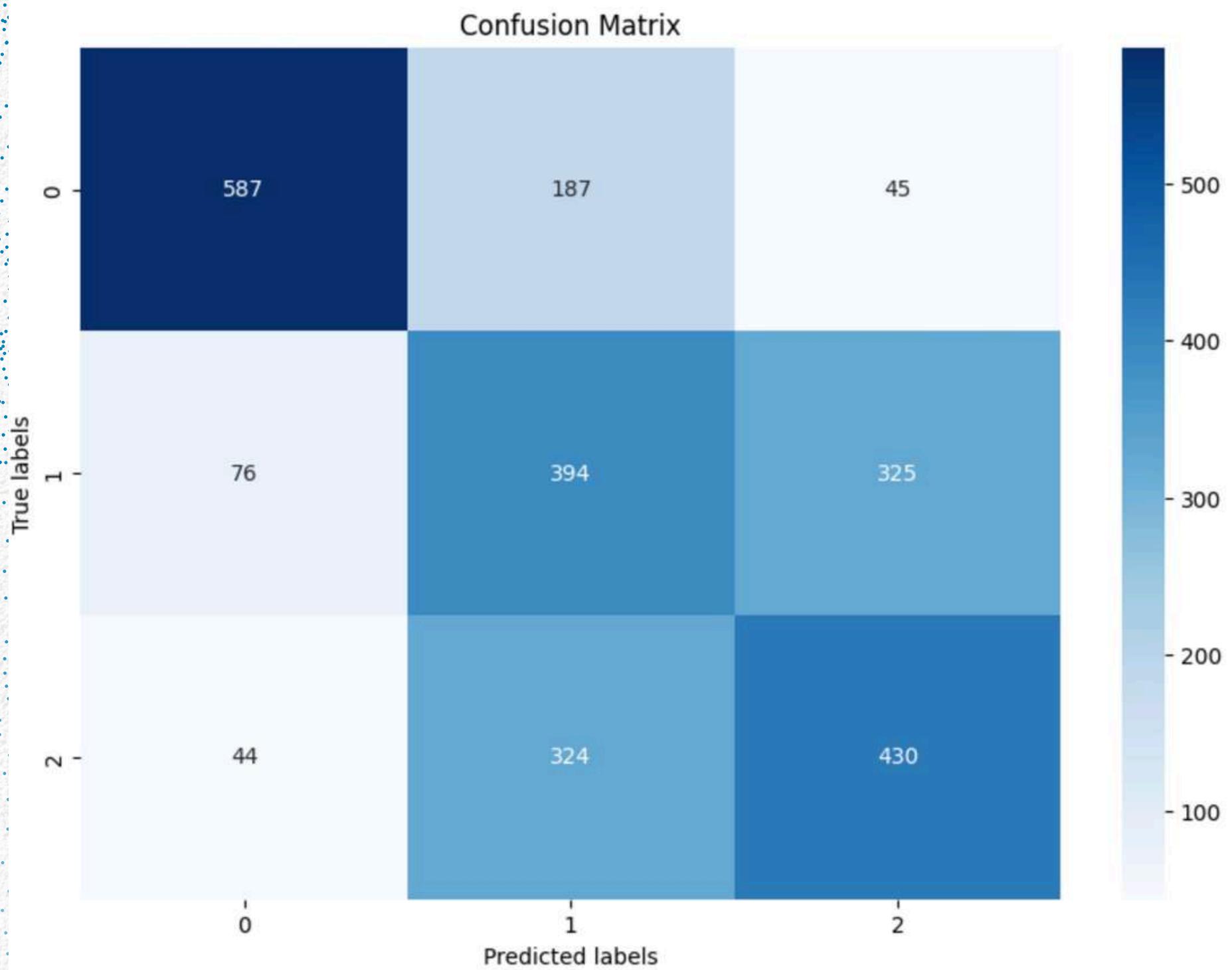
F1-Score: 59,03%

Precision: 60,10%

Recall: 58,37%



Matriz de confusão - Word2Vec + Embedding Layer



Sprint 3

Modelos utilizados

Word2Vec + Naive Bayes

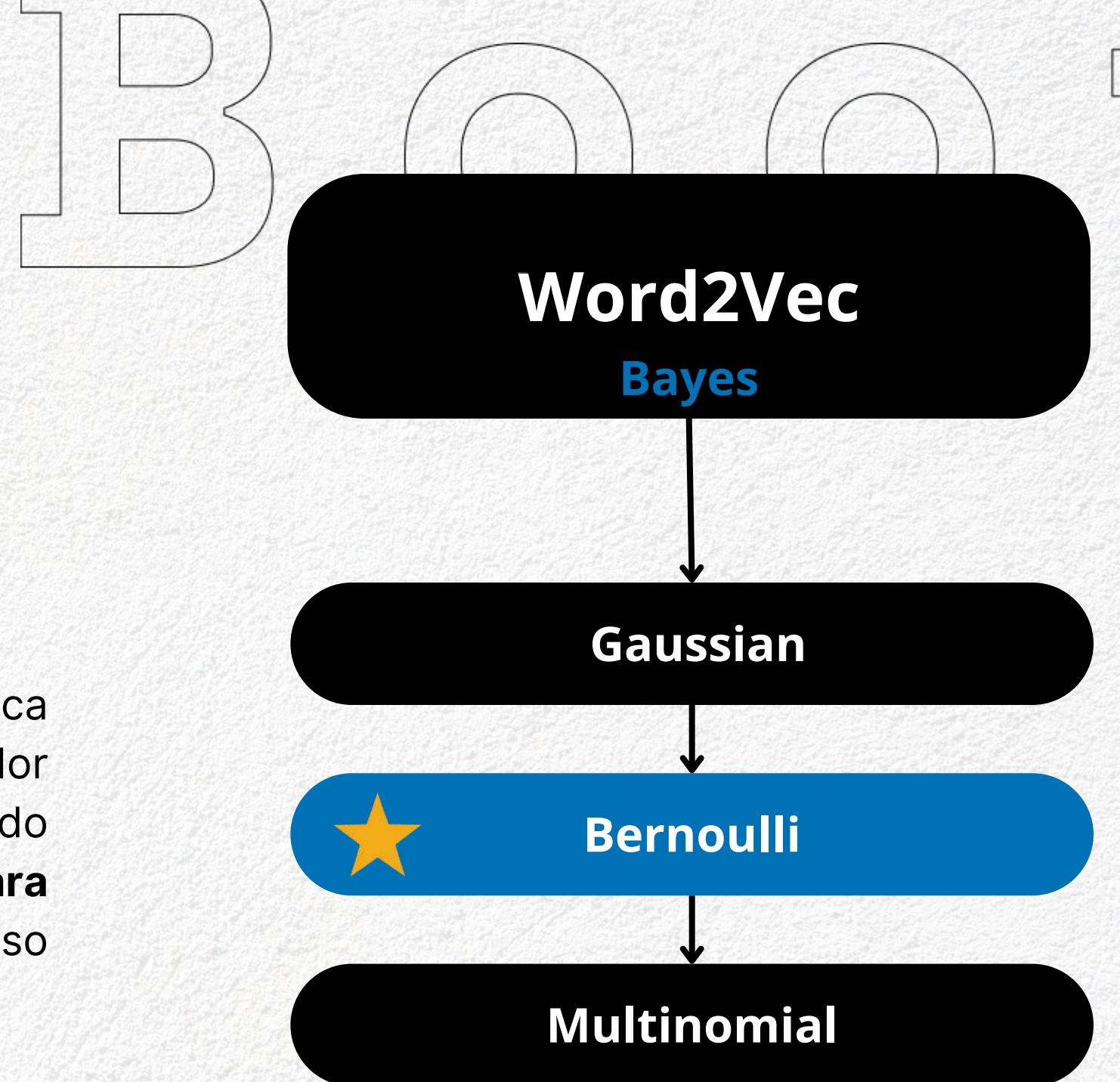
A combinação de Word2Vec e Naive Bayes une a representação semântica eficaz de palavras com a simplicidade e eficiência de um classificador robusto. Word2Vec transforma palavras em vetores densos, capturando significados, enquanto **Naive Bayes utiliza essas características para melhorar a classificação de texto.** Isso resulta em um modelo mais preciso e escalável para tarefas de PLN.

Resultados >>> Bernoulli

Precision: 66,17%

Recall: 61,07%

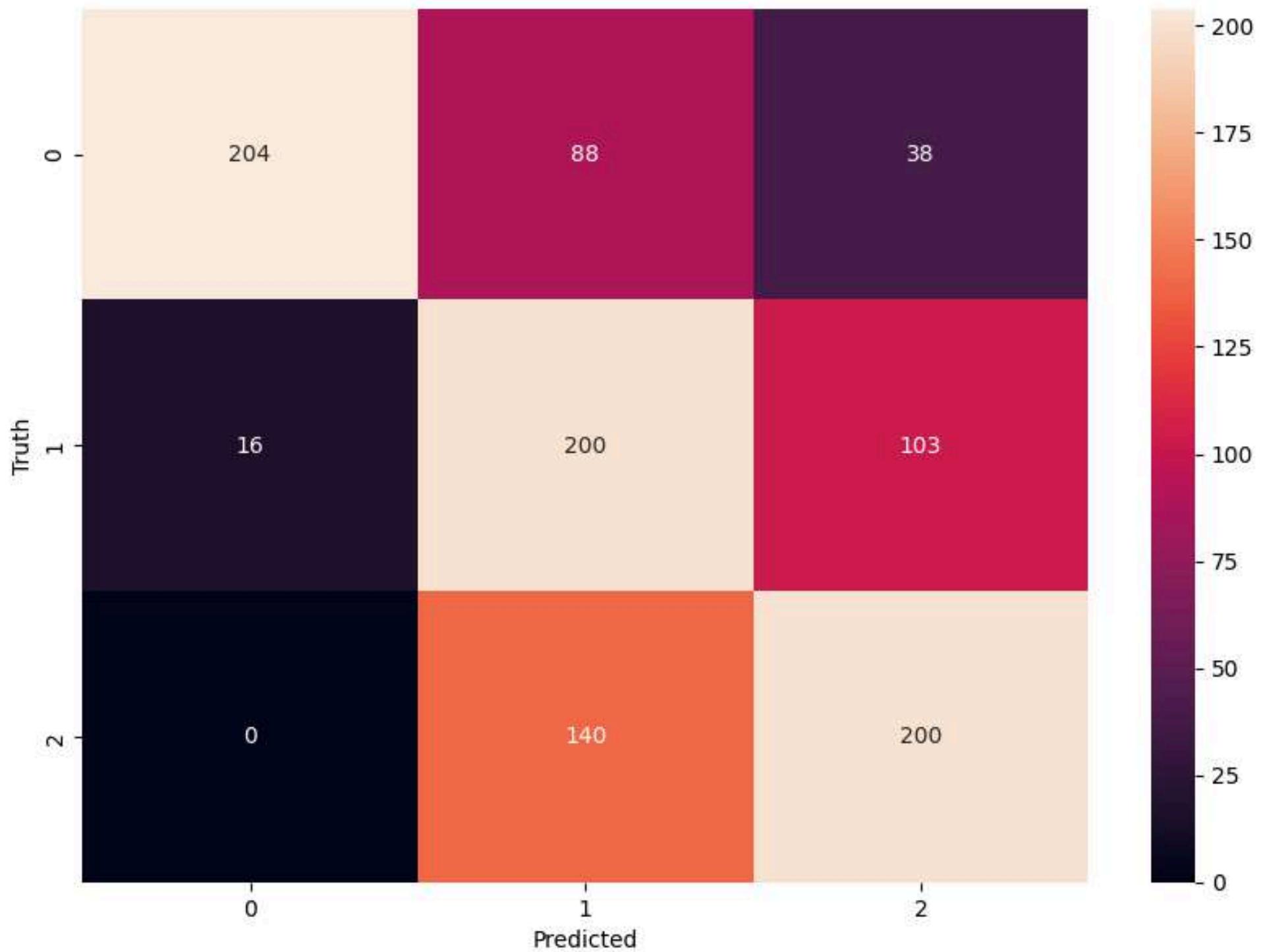
F1-score: 61,23%



Boot



Matriz de confusão - Word2Vec + Naive Bayes (Bernoulli)



FastText

Vantagens

Captura semântica
e morfológica

Adequado para
grandes volumes
de dados

Treinamento rápido

FastText

FastText

FastText + SVM

Bigramas e Softmax

Accuracy: 57,17%
Precision: 75,81%
Recall: 57,17%
F1-Score: 58,08%

Accuracy: 98,78%
Precision: 98,79%
Recall: 98,78%
F1-Score: 98,78%

Accuracy: 99,82%
Precision: 99,82%
Recall: 99,82%
F1-Score: 99,82%

Boot



Overfitting?

Pré-processamento

1. Requisitos Específicos do FastText;
2. Manter caracteres relevantes para subpalavras.

Bigramas

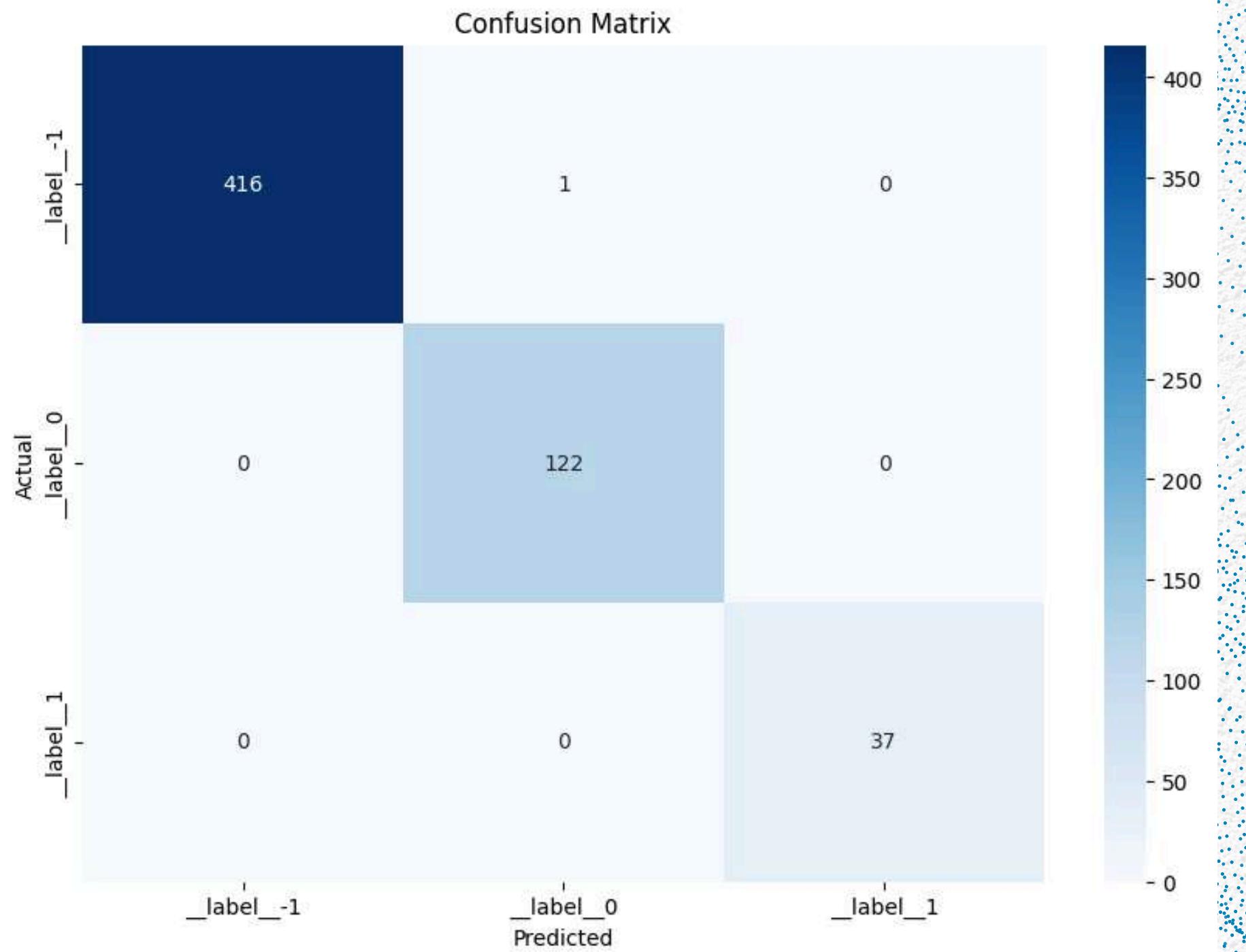
1. Ordem das palavras;
2. Identifica expressões idiomáticas;
3. Desambiguação de significados.

Softmax Hierárquico

1. Treinamento mais rápido;
2. Aplicabilidade em modelos de Embedding;
3. Memória.



Matriz de confusão - FastText + SVM



Cross-validation FastText + SVM

Métricas de desempenho nos dados de teste:

Accuracy: 99,65%

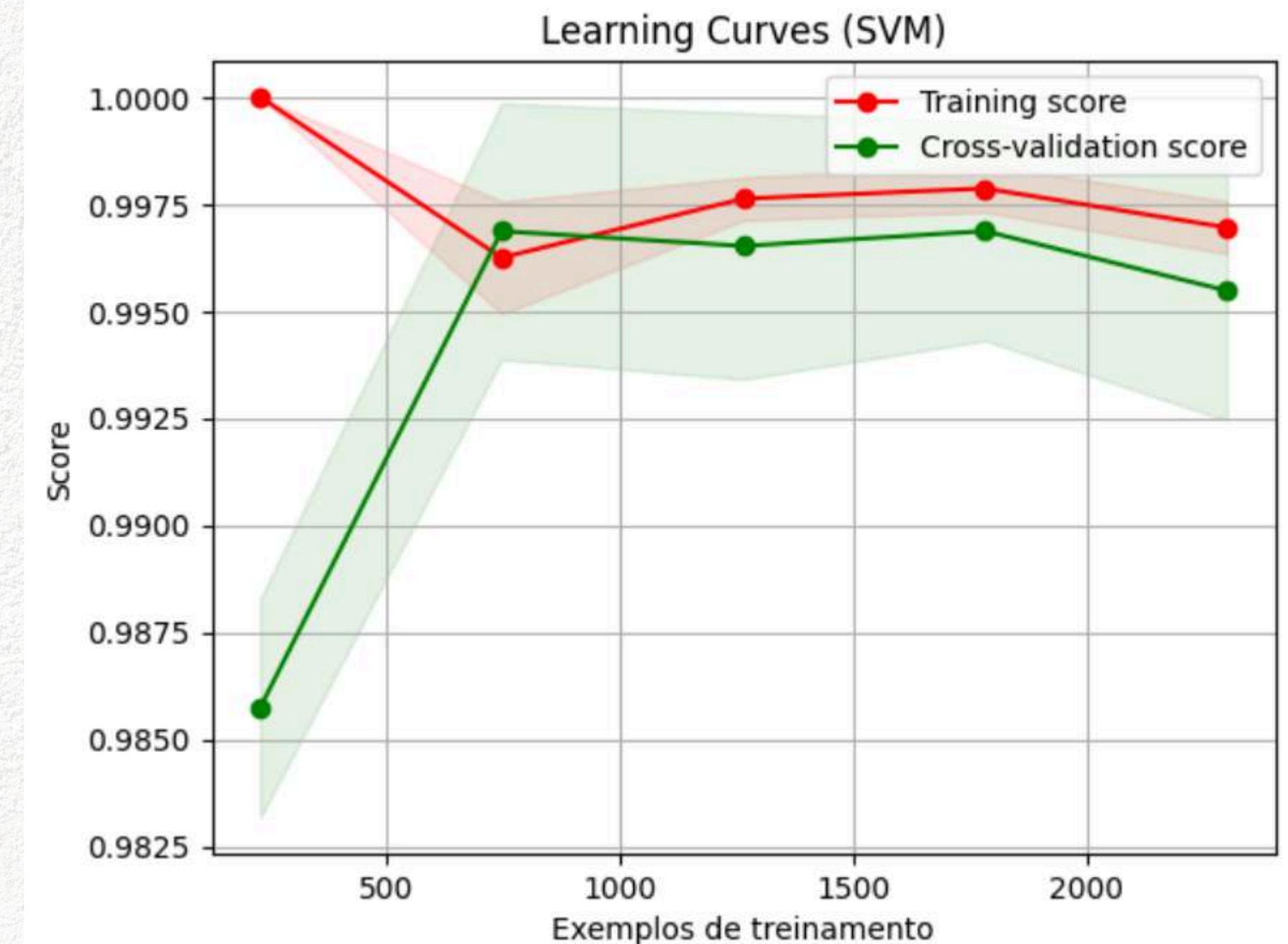
Precision: 99,65%

Recall: 99,65%

F1-Score: 99,65%

Considerações sobre Overfitting

- **Alto Desempenho:** As métricas relatadas mostram que o modelo apresenta um **alto desempenho**, tanto durante a **validação cruzada** quanto nos dados de teste.
- **Próximos passos:**
 - Teste em Novos Dados;
 - Avaliar com Outras Métricas.



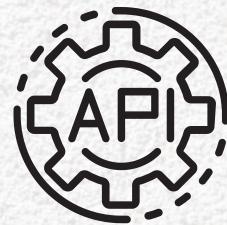
Conclusões da Sprint

Avanços significativos: O modelo **FastText + SVM** apresentou métricas significativamente maior comparado aos da sprint anterior.

Melhorias de desempenho: A integração de **FastText com SVM demonstrou ser uma abordagem altamente eficaz** para o processamento de linguagem natural.



Próximos Passos: Sprint 4



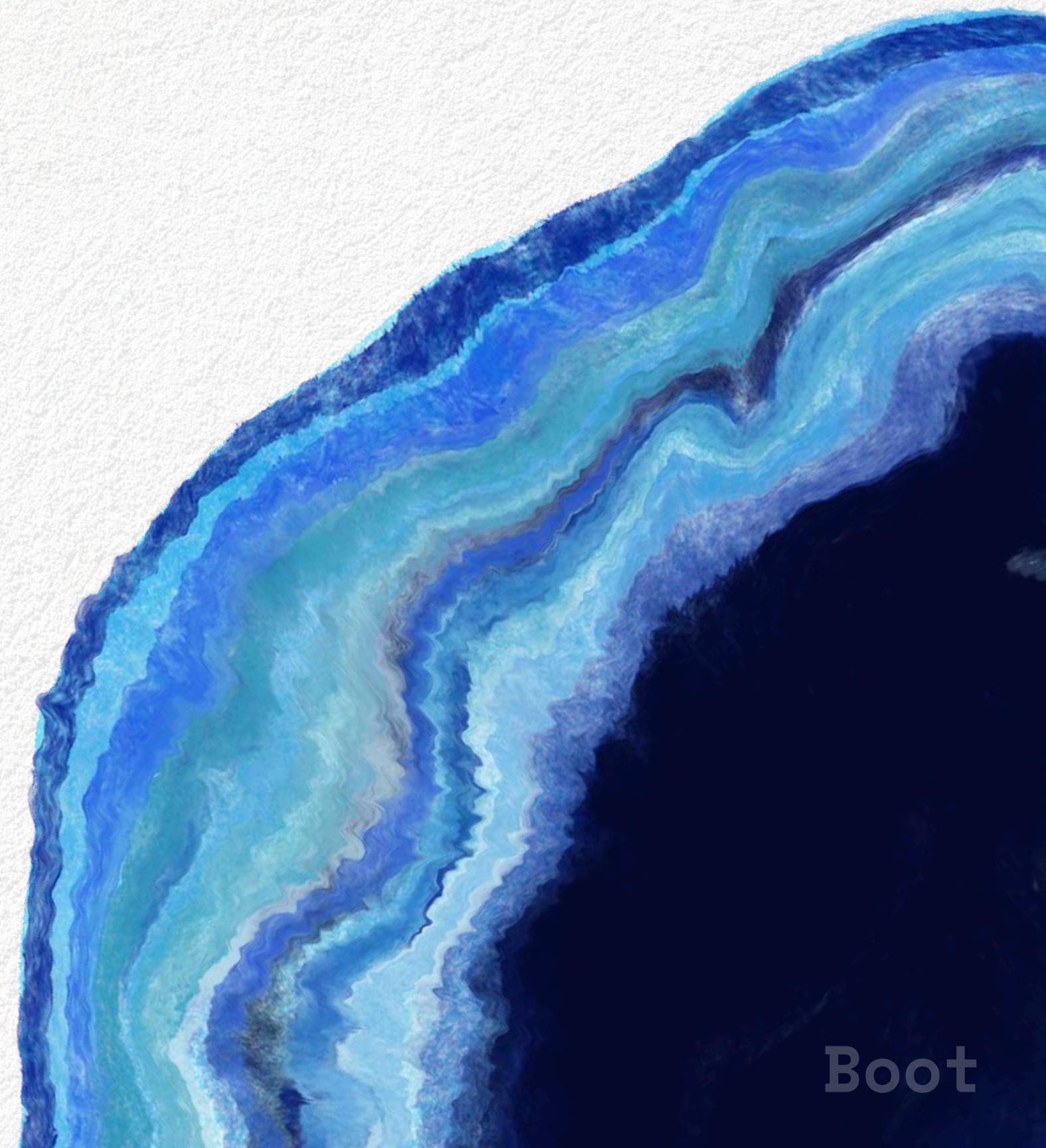
Implementação da API



Testar modelo com novos datasets



Front-end



Boot



Boot



Matheus R.



Henrique Cox



Eduardo HOS



Mateus M.



Marcelo Saasi



Otto



Celine



Boot