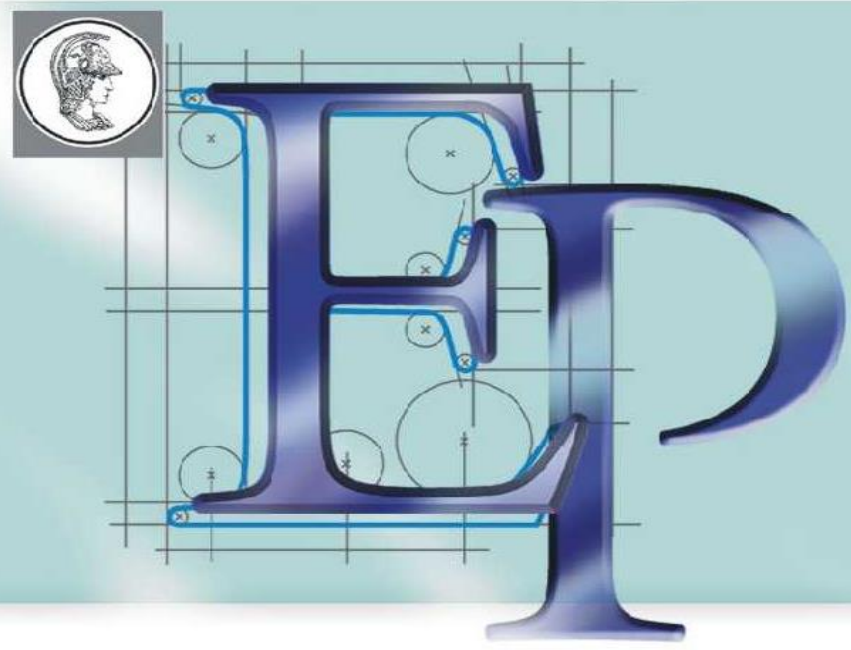


Projeto de Formatura – 2022



PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia de Computação

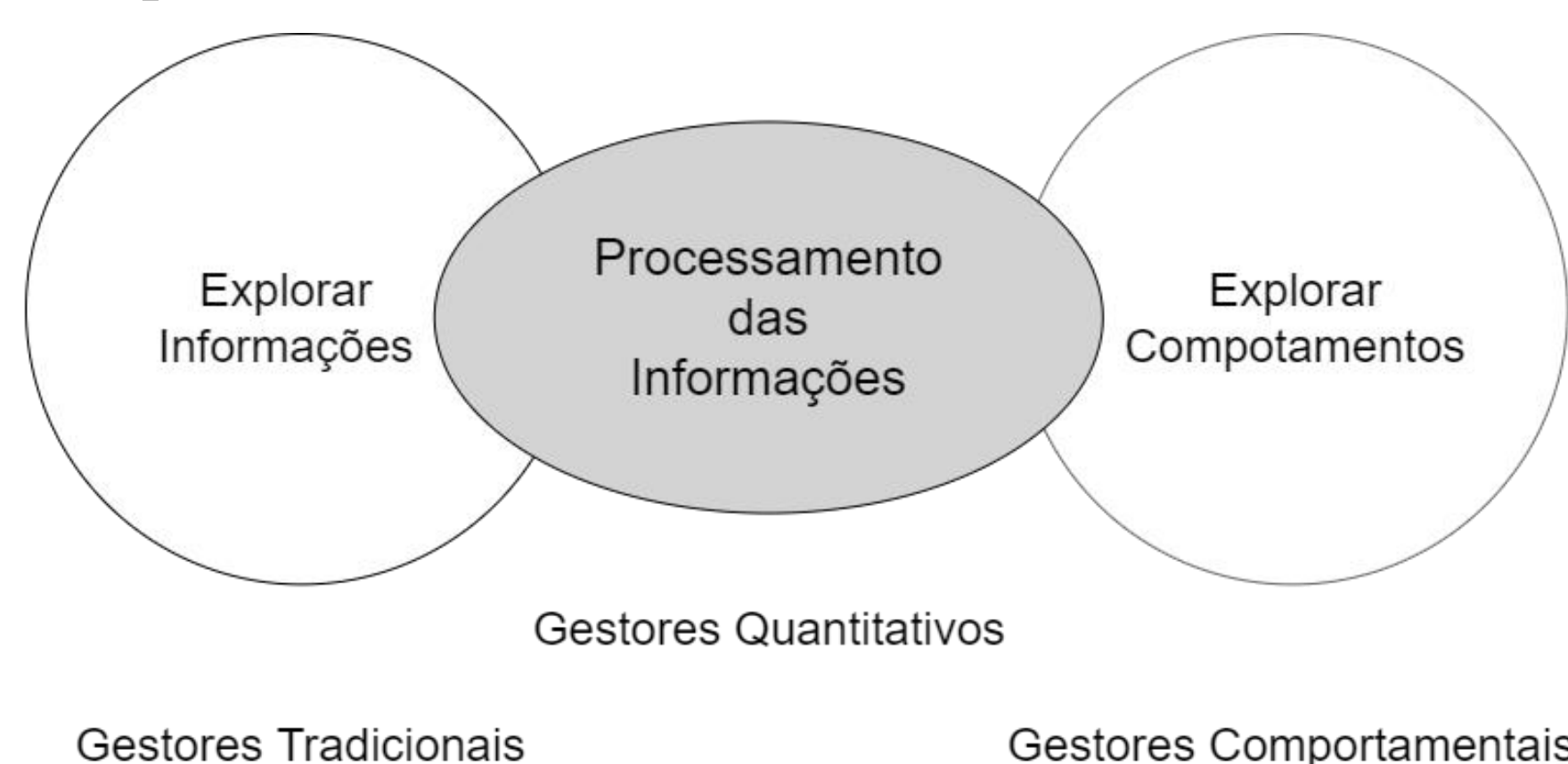
Tema: **FinBERT PT BR: Análise de sentimentos de textos em português referentes ao mercado financeiro**

O que é o FinBERT PT BR?

O FinBERT PT BR é um modelo de redes neurais capaz de estimar o sentimento de textos em português no contexto do mercado financeiro. **Dado um texto ele é capaz de dizer se o sentimento é positivo, neutro ou negativo.** Com esse modelo é possível classificar grandes quantidades de notícias automaticamente.

Por que utilizar IA no Mercado Financeiro?

A inteligência artificial (IA) pode ser utilizada para **aprimorar o processamento de informações públicas**, auxiliando investidores no processo decisório, e por consequência apresentando um retorno dos investimentos acima da média (**alfa**).

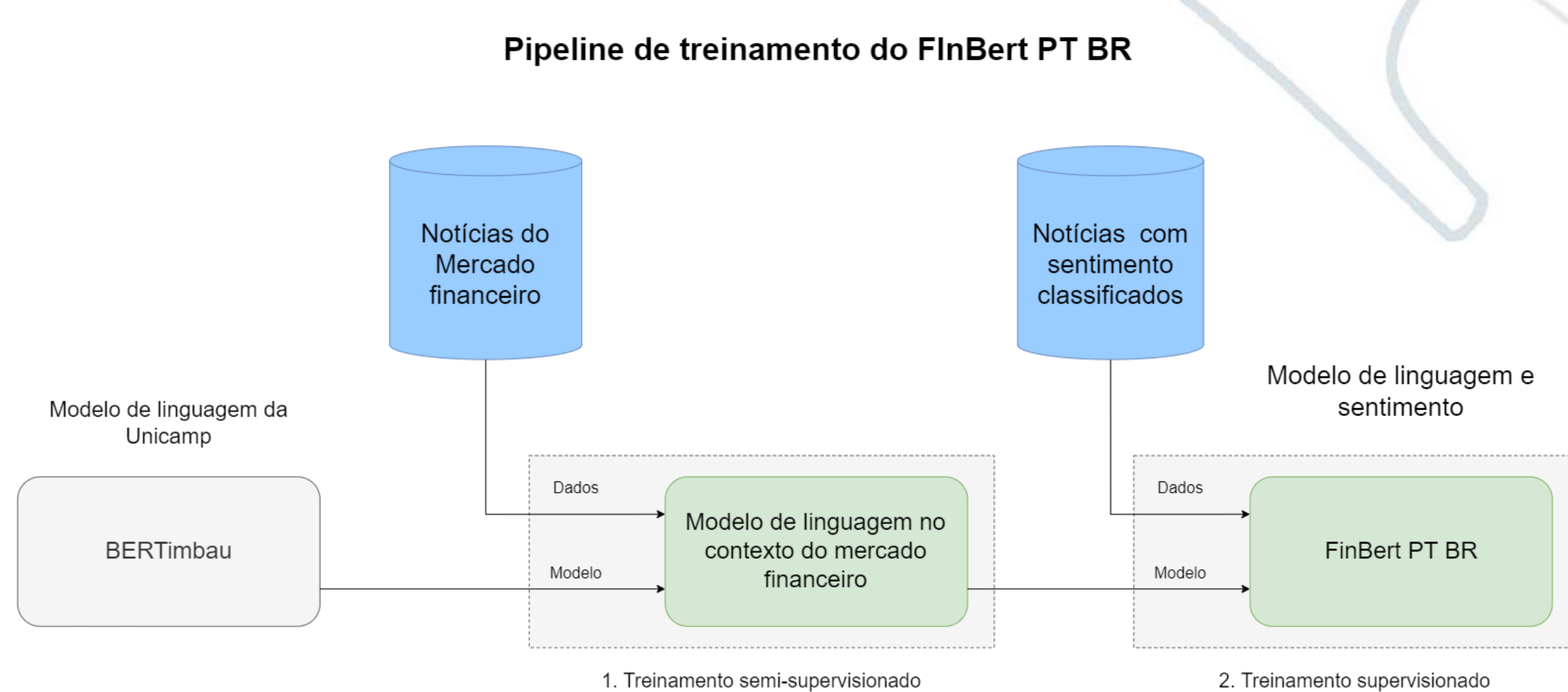


Por que treinar modelos em Português?

Atualmente **existem poucos modelos de processamento de linguagem natural em português**. Em inglês existem mais de 6000 modelos, já em português apenas 200, sendo a maioria deles treinados em múltiplas linguagens.

Arquitetura do projeto

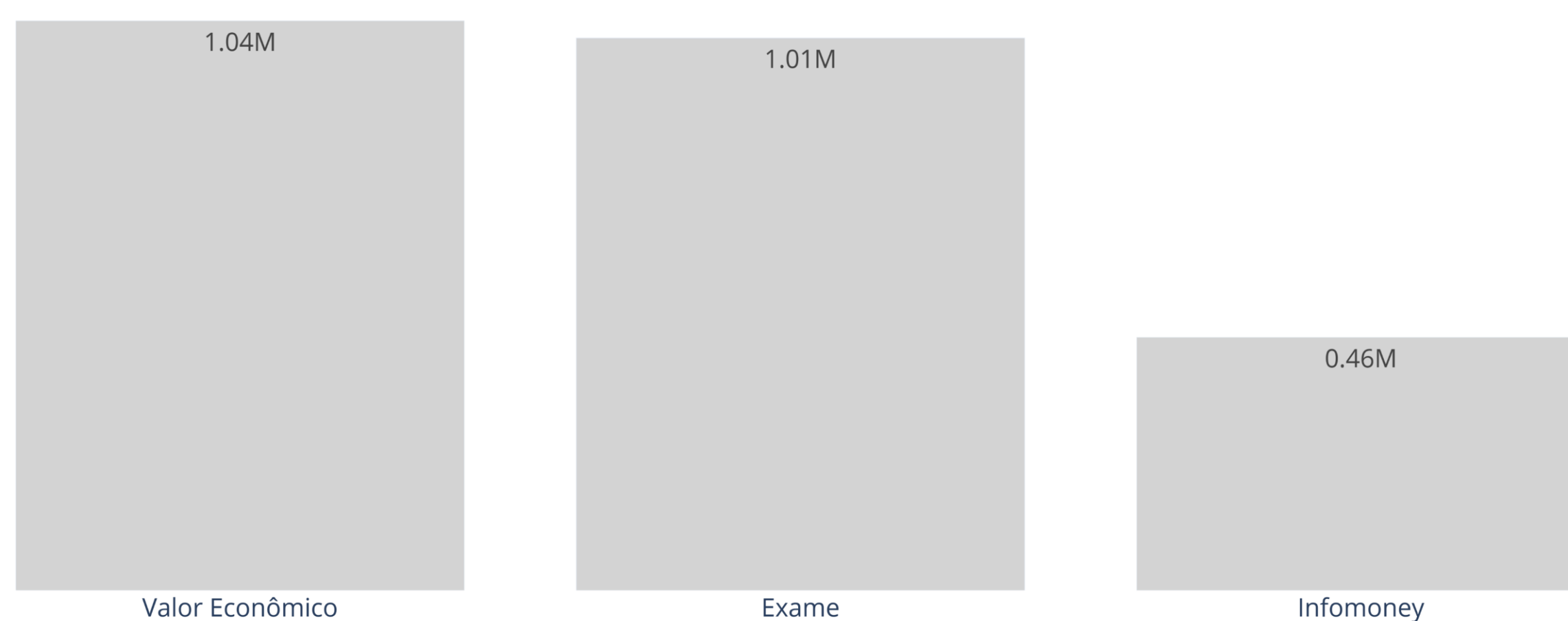
A arquitetura da solução é baseada em um modelo de linguagem genérico (BERTimbau), que é especializado em textos do mercado financeiro e logo depois treinado para a tarefa de análise de sentimentos a partir de dados rotulados manualmente. Portanto, o modelo é **especializado com a semântica do mercado financeiro e é capaz de estimar o sentimento de textos de notícias**.



Aquisição de dados

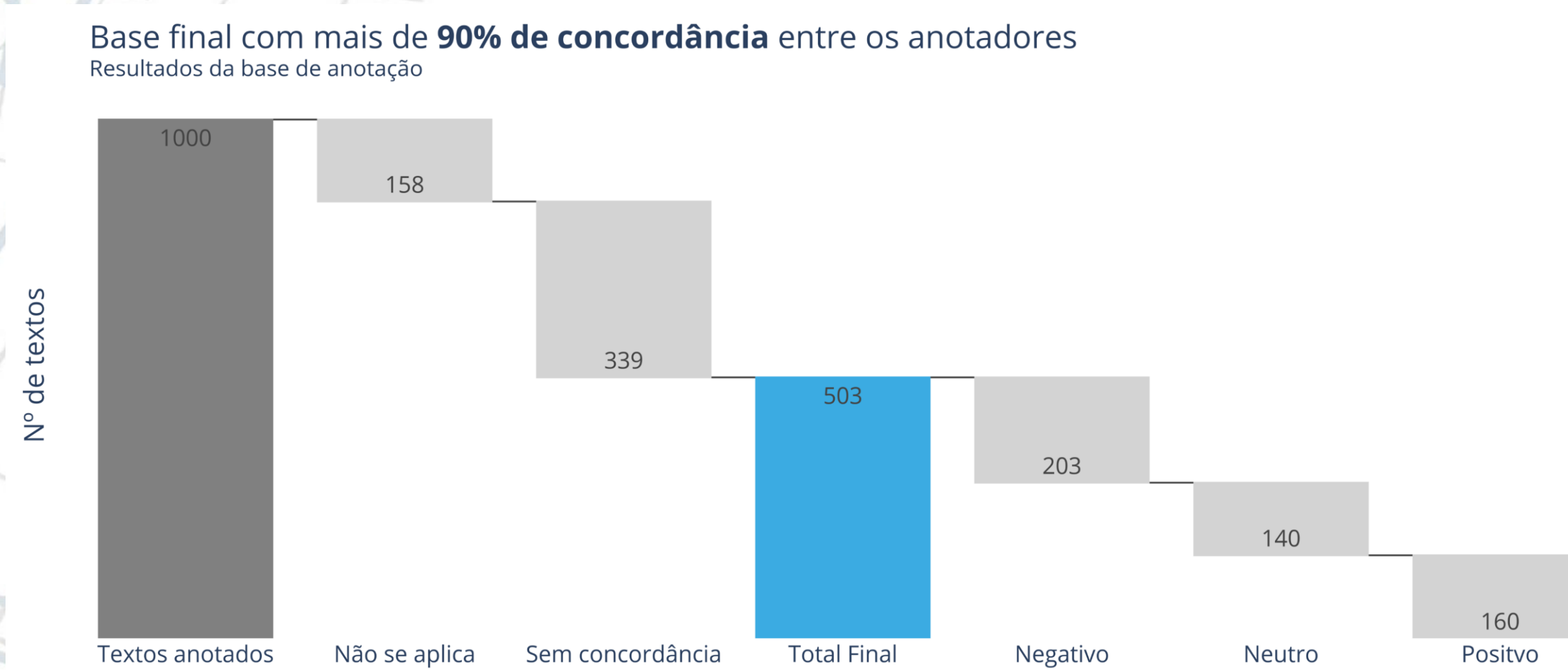
Para treinar os modelos foi necessário adquirir dados textuais de portais de notícias do mercado financeiro. Os portais escolhidos foram: Valor Econômico, Infomoney e Exame; E no total foram coletados mais de **2,5 milhões de textos**.

2.51M de textos no contexto do mercado financeiro coletados, totalizando **120.0M de palavras**
Número de total de textos coletados (títulos, subtítulos e parágrafos) por fonte de dados



Anotação de sentimentos

Para treinar o modelo de sentimentos foi necessário construir uma base de textos com os sentimentos rotulados. O processo de anotação foi realizado por 3 pessoas para evitar vieses, e no total foram **anotados 1000 textos**:



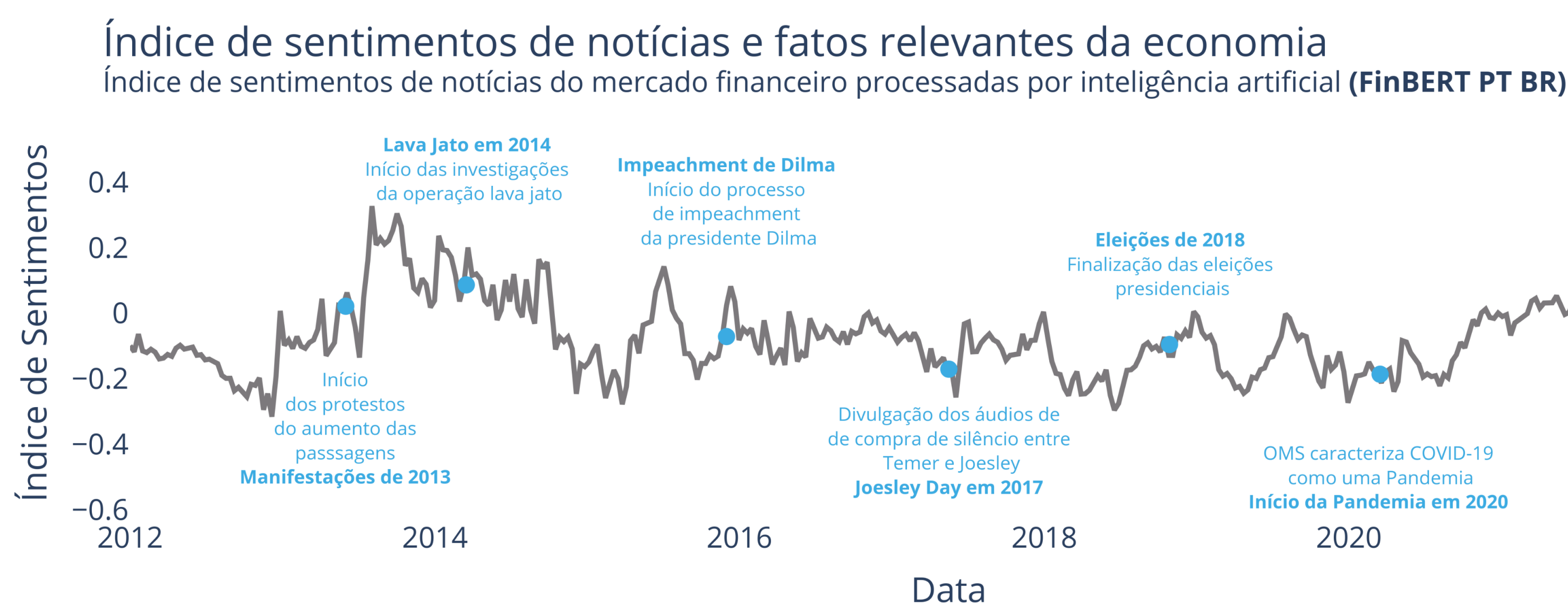
Modelo de sentimentos

O modelo de sentimentos apresentou **métricas melhores que os principais modelos no estado da arte**

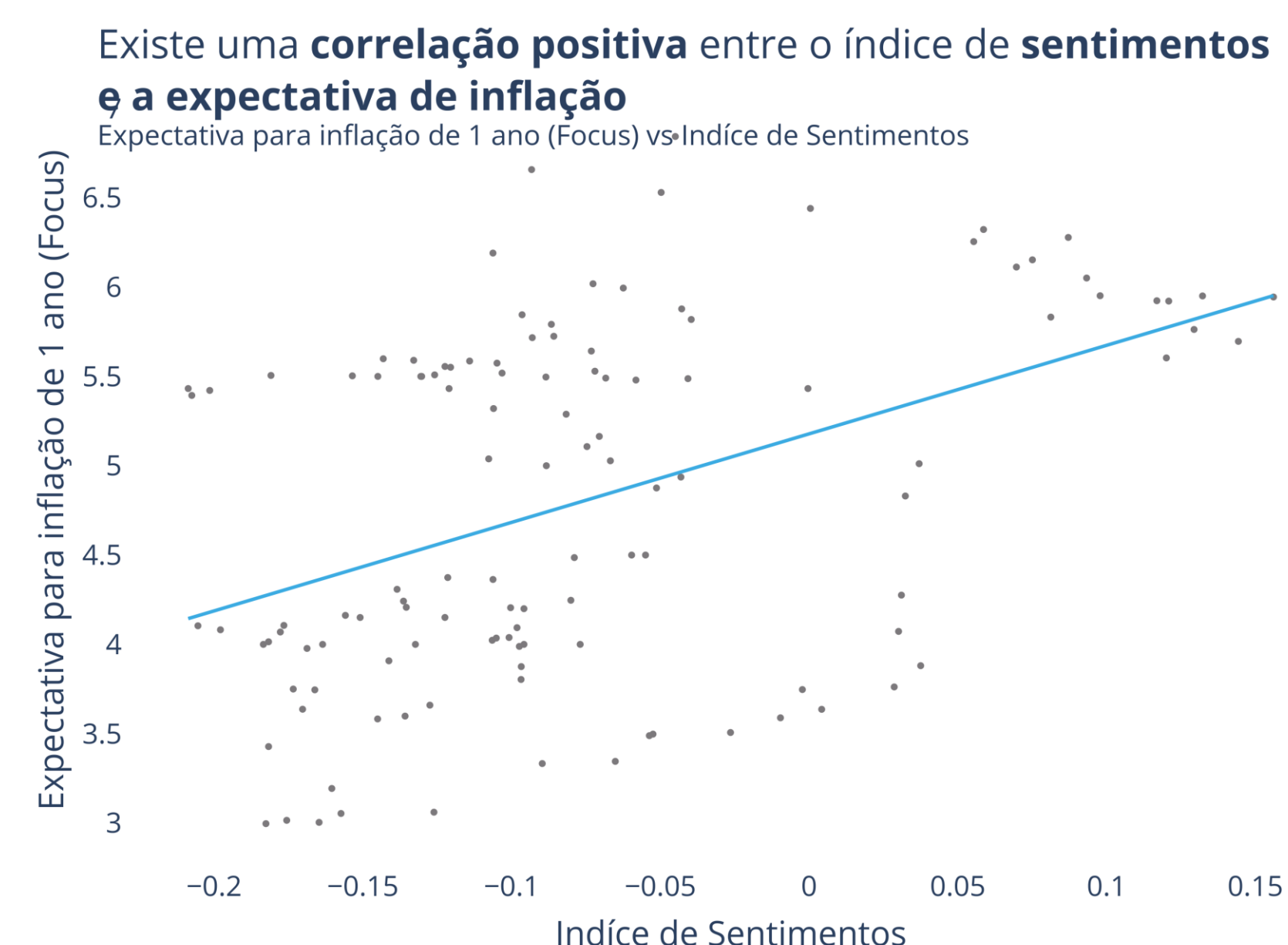
Modelo	Acurácia	F1-Score
Random Forest com TFIDF	45%	0.35
FinBERT (EN)	67%	0.67
BERTimbau	69%	0.63
FinBERT PT BR	76%	0.73

Aplicações do FinBERT PT BR

Construção de índices de sentimentos:



Análise de expectativa de inflação:



Integrante: Lucas Leme Santos

Professor(a) Orientador(a): Profa. Dra. Anna Helena Reali Costa