



From Chaos to Insight

Automatic proposal classification and
recommendation on Brasil Participativo with
BERTopic

2025





- **Federal Government platform** designed to expand dialogue with society and strengthen participatory democracy.
- Launched in 2023, it enables citizen participation in public policies through Public Consultations, National Conferences, Participatory Plans, as well as Councils and Forums.

62 processes

38 assemblies

+1,6 M users

Proposals in numbers

11
processes

13
components

24.344
proposal

1.870.691
votes

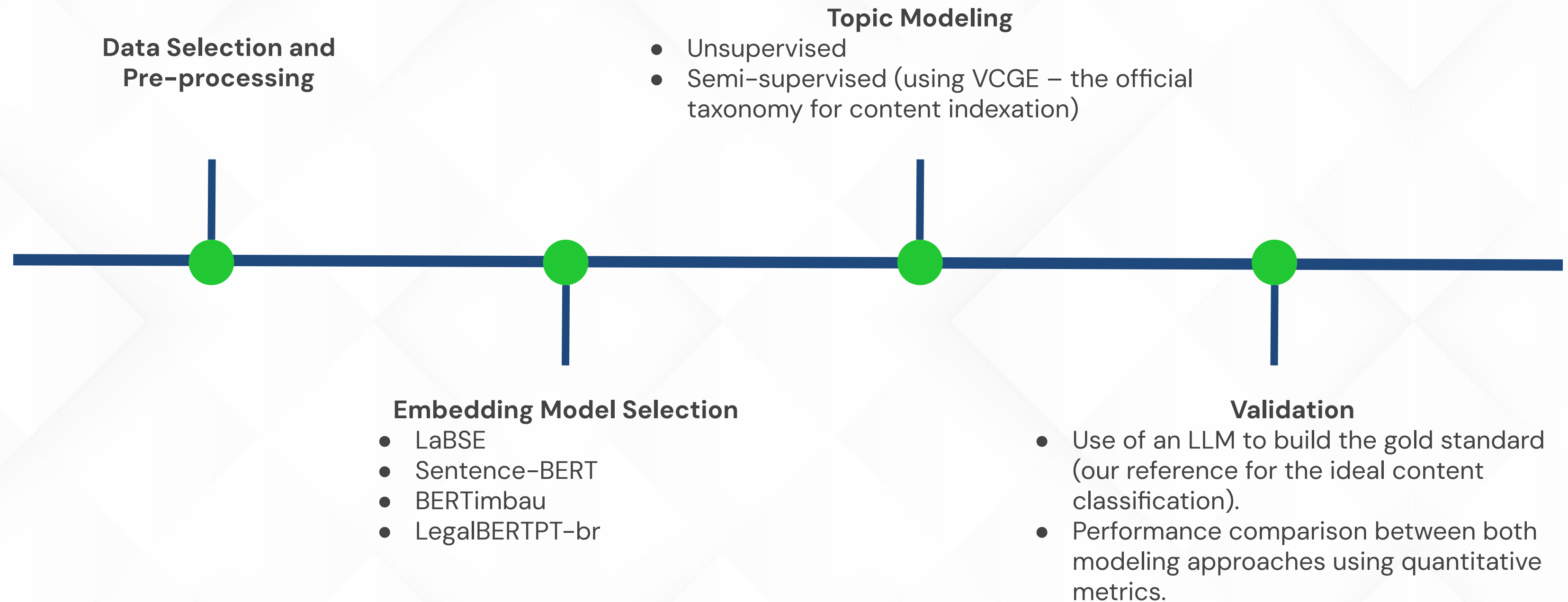
1.246.519
users voting only once

~80%
vote only once

How can we automatically identify themes in citizen proposals and use it to improve engagement?

We explored **BERTopic**-based models and both **unsupervised** and **semi-supervised** approaches to develop our final pipeline.

Path to build the pipeline

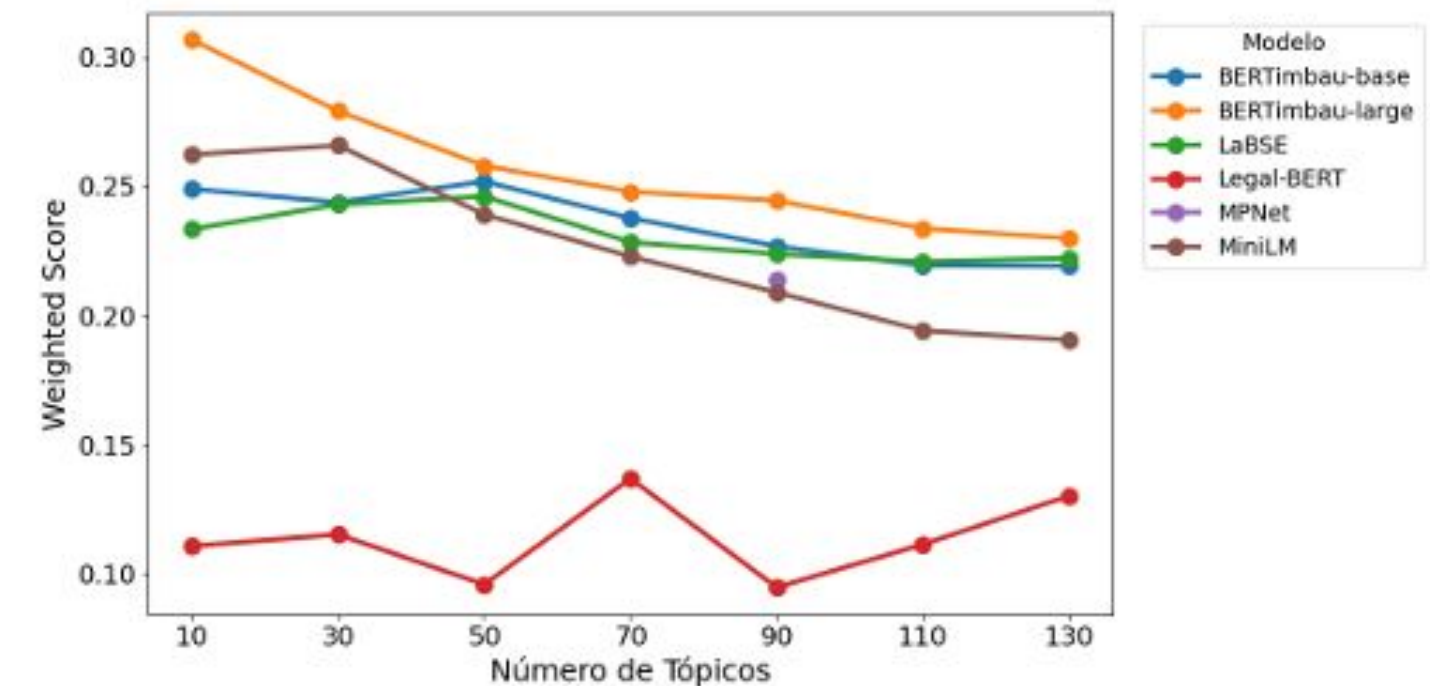


Data Selection and Pre-processing

- 10,186 proposals collected from:
 - Plano Clima
 - PPA Participativo
 - Congresso da Juventude
- Corpus split: 80% training / 20% testing.
- Pre-processing steps:
 - Lowercasing all text
 - Removing punctuation and stopwords
 - Lemmatization (e.g., “cidadãos” → “cidadão”)

BERTimbau-large was chosen because it best captured the semantic meaning of Brazilian Portuguese and produced coherent, representative topics among all models tested.

Embedding Model Selection



- Normalized Coherence (NC): How much the words within a topic make sense together.
- Normalized Diversity (ND) How distinct the topics are from each other. Higher is better.
- Weighted Score (WS): $WS = 0.8 \times NC + 0.2 \times ND$

Topic modeling

VCGE

- Unsupervised approach:
 - BERTopic ran freely, discovering topics only from text patterns.
 - Captured spontaneous citizen themes but lacked institutional alignment.
- Semi-supervised approach:
 - Introduced seed words based on VCGE categories.
 - Guided topic formation toward official government taxonomy.
- Evaluation metrics:
 - Compared both using Coherence (NC), Diversity (ND), and Alignment (ARI/NMI).

Tabela 1. Mapeamento resumido de 12 categorias do VCGE (versão 2.1.0)

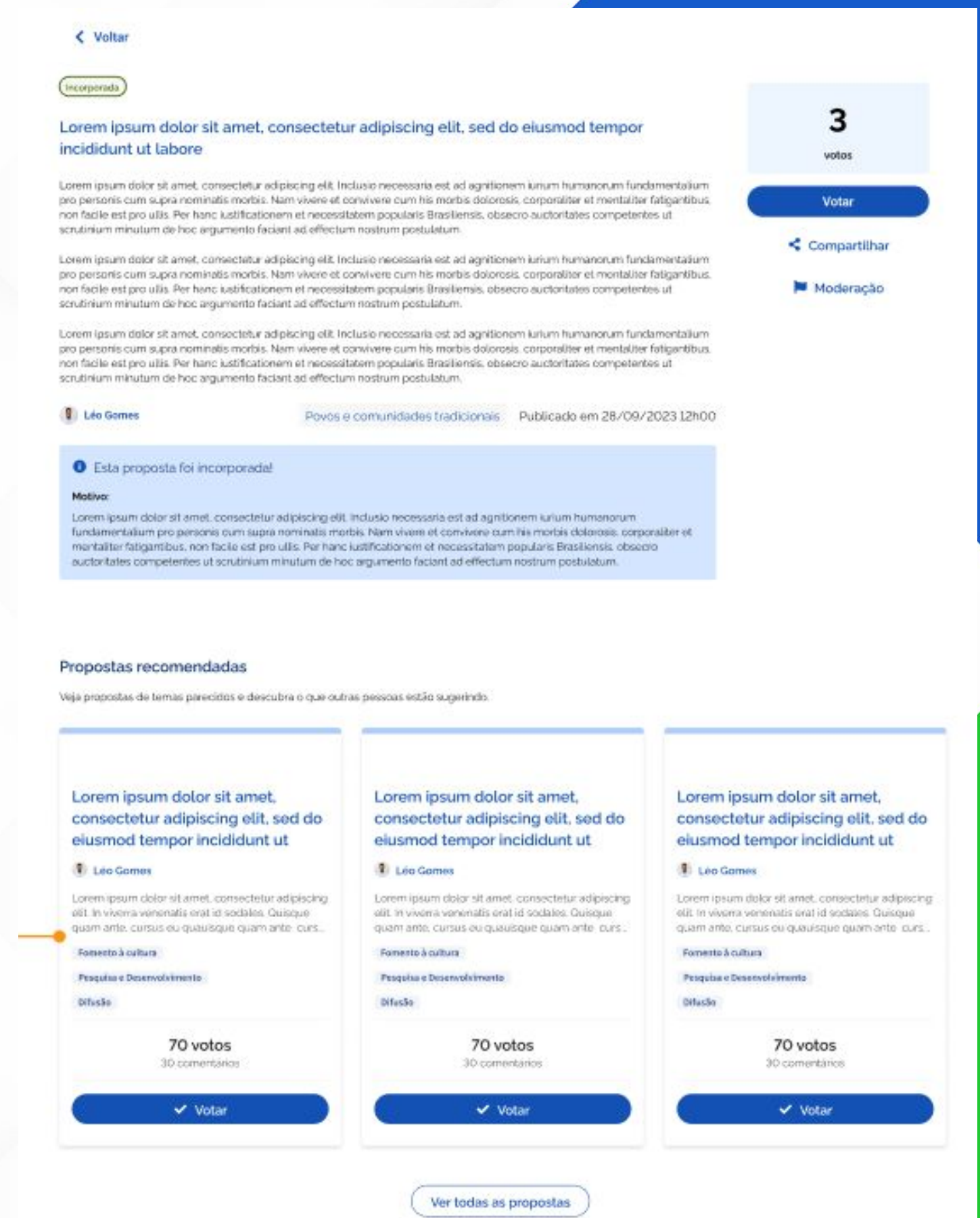
| Nível 1 (N1) | Nível 2 (N2) |
|----------------------------|--------------------------------------------------------------------------------|
| Agropecuária e Pesca | Defesa e vigilância sanitária, Produção agropecuária, Outros em Agropecuária |
| Comércio e Serviços | Comércio externo, Defesa do Consumidor, Turismo, Outros em Comércio e Serviços |
| Comunicações | Comunicações Postais, Telecomunicações, Outros em Comunicações |
| Cultura | Difusão Cultural, Patrimônio Cultural, Outros em Cultura |
| Defesa Nacional | Defesa Civil, Defesa Militar, Outros em Defesa Nacional |
| Energia | Combustíveis, Energia elétrica, Outros em Energia |
| Esporte e Lazer | Esporte comunitário, Esporte profissional, Lazer, Outros em Esporte e Lazer |
| Habitação | Habitação Rural, Habitação Urbana, Outros em Habitação |
| Indústria | Mineração, Produção Industrial, Propriedade Industrial, Outros em Indústria |
| Pesquisa e Desenvolvimento | Difusão, Outros em Pesquisa e Desenvolvimento |
| Saneamento | Saneamento Básico Rural, Saneamento Básico Urbano, Outros em Saneamento |
| Urbanismo | Infraestrutura Urbana, Serviços Urbanos, Outros em Urbanismo |

VCGE is the Brazilian government’s official controlled vocabulary

- Standardized taxonomy used to classify and organize public information across agencies.
- Structured in two levels: **Level 1 (N1)** represents broad thematic domains, and **Level 2 (N2)** contains their specific subcategories.

Discussion and Application

- Proposal recommendations for voting and public consultation.
- Automated reports for government agencies and citizens.
- Chatbot integration, using categorized knowledge to improve user interaction and access to information.
- The pipeline can be replicated in federal, state, or municipal platforms and continuously improved through expert feedback.





Thank you!

Lab Livre UnB: @lab.livre

Bruna Pinos: bruna.pinos@gmail.com

2025

