

Soluções para a Integração e Gestão da Mobilidade Aérea

Integração da mobilidade aérea avançada à mobilidade urbana e no desenvolvimento sustentável das cidades





WP 1 Pipeline - Pontuação Pestel

Processo de Embeddings

• Embeddings:

- Definição das Categorias
- Le os arquivos
- Transforma em Vetoriais
- Compara Ambos

✓ output

≡ analyse_pestel_relatorio.txt

📄 analyse_pestel.csv

{ } analyse_pestel.json

📊 analyse_pestel.xlsx

```
Projeto > pestel.py > PESTELAnalyzer > setup_pestel_definitions
25 class PESTELAnalyzer:
61     def setup_pestel_definitions(self):
62         """
63         Define as categorias PESTEL em português e inglês para maior robustez.
64         """
65         # Definições em português (baseadas no documento original)
66         self.pestel_definitions_pt = {
67             "Político": """
68             Referem-se à influência do governo, à formulação de políticas públicas e à estabilidade política que fundamentam o ecossistema c
69             """,
70             "Econômico": """
71             Referem-se à viabilidade financeira, sustentabilidade dos modelos de negócio e impacto macroeconômico do ecossistema da Mobilida
72             """,
73             "Social": """
74             Referem-se às tendências demográficas, culturais, comportamentais e de percepção da sociedade que impactam a adoção e a integraç
75             """,
76             "Tecnológico": """
77             Referem-se ao desenvolvimento, integração e maturidade do ecossistema de tecnologias de ponta que viabilizam as operações da Mo
78             """,
79             "Ambiental": """
80             Referem-se aos impactos ecológicos, à conformidade com a legislação verde e à gestão de recursos naturais que fundamentam a sust
81             """,
82             "Legal": """
83             Referem-se ao conjunto de leis, regulamentos, normas e frameworks jurídicos que regem e viabilizam as operações seguras e ordena
84             """,
85         }
86     }
87     }
88     }
89     }
90     }
91     }
```



Resultados Obtidos

ICA 100-40_06.06.23.pdf

PT

Tamanho do texto: 110488 caracteres

1.7

Ambiental

5.5

Econômico

10.0

Legal

8.7

Político

1.1

Social

10.0

Tecnológico

Lei_PNMU.pdf

PT

Tamanho do texto: 28215 caracteres

1.2

Ambiental

5.8

Econômico

10.0

Legal

9.5

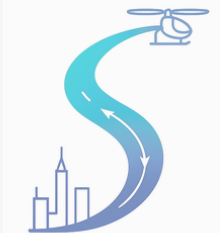
Político

7.0

Social

3.8

Tecnológico



Vantagens

- Captura significado, não só palavras
 - Ex.: “administração pública” $\approx$ “política”, mesmo sem usar a palavra exata.
- Escalável
 - Processa milhares de documentos rápido, pois embeddings são vetores.
- Explicável e ajustável
 - Boost de palavras-chave melhora transparência e controle.

Limitações

- Perda de contexto longo
 - Embedding único não capta bem textos muito grandes.
- Modelo não treinado para PESTEL
 - Modelo usado é geral, não jurídico $\rightarrow$ pode faltar precisão.
- Pontuação relativa
 - Similaridade $\neq$ probabilidade.
 - Ex.: 8/10 = mais parecido com Político do que com Tecnológico.



- **Origem e Arquitetura**

- Criado pela Microsoft Research (Wang et al., 2020).
- Versão compacta do BERT/RoBERTa com distillation.

- **Treinamento**

- Corpus multilíngue (Wikipedia, livros, CommonCrawl).
- Versão **paraphrase-multilingual-MiniLM-L12-v2** ajustada para embeddings semânticos em 50+ idiomas.

- **Dimensão e Eficiência**

- 12 camadas, ~33M parâmetros (BERT-base tem 110M).
- Embeddings de 384 dimensões.

- **Aplicações**

- Similaridade semântica, classificação, busca, análise PESTEL.

- **Fonte**

- Disponível no HuggingFace: [sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2](https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2).

Uso do Programa

- Divergia quanto a buscas de Leis .
- Não tivemos sucessos quanto a integração ao código.
- **Mais teste são necessários para avaliar melhor a ferramenta.**



☒ Tudo ☐ Legislação ☐ Jurisprudência ☐ Proposições Legislativas ☐ Doutrina

[Pesquisa Avançada](#) | [Acervo](#) | [Sobre o LexML](#) ([English](#), [Français](#), [Español](#)) | [FAQ](#) | [Manual](#) **(Novo!)**