

# Análise das Categorias dos Laudos

## Informações importantes para classificação (Normal/Anormal)

### Dados rotulados/classificados:

Os dados utilizados para análise estão disponíveis em:  
<https://docs.google.com/spreadsheets/d/1aFX307CvFUIsfeKFzKORvnx1yl6neVpEnZCF7wsQ-mI/edit?usp=sharing>

## Notebook: categoriasLaudos.ipynb

Este notebook realiza a análise das categorias dos laudos médicos a partir dos dados rotulados.

### Etapas principais do notebook

- **Leitura dos dados:** Importa o arquivo .csv contendo os laudos já rotulados.

### Definição das categorias

#### Categorias secundárias:

```
CATEGORIAS_SECUNDARIA_LAUDOS = ['Vascular', 'Inflamacao', 'Degeneracao', 'Normal']
```

#### Categorias adicionais:

```
CATEGORIAS_ADICIONAIS_LAUDOS = ['Inflamacao', 'Morfologia', 'Trauma', 'Metabolico', 'Degeneracao', 'Artefato', 'Vascular']
```

### Função make\_prompt

Gera um prompt formatado a partir de uma pergunta (**query**) e de um texto de contexto (**relevant\_passage**). Este prompt é enviado ao modelo de IA generativa Gemini para obter respostas.

### Função descricao\_rotulos\_laudos

Utiliza a **make\_prompt** para interagir com o modelo e gerar insights sobre os laudos, como os bigramas e trigramas mais relevantes para cada categoria de diagnóstico.

## Exemplo de aplicação

Para cada categoria, é feita uma consulta como:

query = "Quais são os bigramas e trigramas mais relevantes no texto abaixo que levam ao diagnóstico: "

campo\_laudo = "Impressão"

descricao\_rotulos\_laudos(CATEGORIAS\_SECUNDARIA\_LAUDOS,  
ROTULO\_SECUNDARIO, query, campo\_laudo, laudos\_classificados)

**Observação importante:** Para todas as categorias, a função retorna os bigramas e trigramas relevantes.

Para as categorias "Inflamação" e "Normal", o modelo pode não conseguir identificar padrões significativos.

### Exemplo para a categoria Normal

- **Categoria:** Normal
- O texto não fornece informações suficientes para determinar os bigramas e trigramas mais relevantes para um diagnóstico "Normal".
- O texto descreve a presença de "focos arredondados de anormalidade de sinal" no cérebro, que pode ser um sinal de migrânea.
- Portanto, o texto não suporta a conclusão de um diagnóstico "Normal" e, consequentemente, não é possível identificar bigramas e trigramas relevantes para esse diagnóstico.

Para queries sobre frases mais relevantes:

query = "Quais são as frases mais relevantes no texto abaixo para o diagnostico: "

descricao\_rotulos\_laudos(CATEGORIAS\_ADICIONAIS\_LAUDOS,  
ROTULO\_ADICIONAL, query, campo\_laudo, laudos\_classificados)

### Retorno exemplo para a categoria Normal:

- **Categoria:** Normal

- Frases mais relevantes no texto:
  - "Restante do exame sem anormalidades evidentes detectáveis pelo método."
  - "sendo este achado frequente em pacientes portadores de migrânea, na dependência de dados clínicos."
- A primeira frase indica que a maioria do exame está dentro dos parâmetros normais. A segunda frase, apesar de mencionar anormalidades, as classifica como "frequentes em pacientes portadores de migrânea", sugerindo que elas podem ser um achado normal para esse tipo de paciente.

## Notebook: juncao\_frases\_relevantes.ipynb

Os dados utilizados para análise estão disponíveis em:  
<https://docs.google.com/document/d/1RMDNUxNnVaBA00dyCeQdwUCOD26EUOo0-jLrSF-Ds3w/edit?usp=sharing>

- Mapeamento das categorias antigas para as novas (explicadas na planilha acima):

### Frases relevantes para exames normais:

normal = [

"Restante do exame sem anormalidades evidentes detectáveis pelo método.",

"ausência de alterações",

"ausência de alteração",

"sem alterações",

"sem alteração",

"ausência alterações estruturais significado patológico",

"ausência alterações significado patológico detectáveis",

"ausência alterações significado patológico identificáveis",

"ausência alterações estruturais encefálicas significado patológico detectáveis",

"ausência alterações estruturais encefálicas",

"sem alterações estruturais encefálicas de significado patológico detectáveis",

"sem anormalidades evidentes"

]

- Cria uma nova tabela .csv após identificar os dados normais, (aqueles que contêm alguma das frases relevantes para normal), adicionando uma coluna “Osborn”, que pode conter “normal” ou NaN.
- Para os casos onde `Osborn = normal` e `Fazekas` está vazio (NaN), atribui-se `Fazekas = 0.0`:

```
df.loc[(df['Osborn'] == 'normal') & (df['Fazekas'].isna()), 'Fazekas'] = 0.0
```

```
df.to_csv('laudos_classificados_com_fazekas_0_classificados.csv', index=False)
```