

Data: 10/11/2023 – Eduardo Corrêa

## Bases de Dados e Tasks Relacionadas a Classificação de Técnicas de Persuasão nas Últimas Edições do SemEval

Ano	Base de Dados	Tarefa(s) de Classificação
2019 e 2021	<b>PTC - textos de artigos</b> <ul style="list-style-type: none"><li>18 classes</li><li>Inglês</li><li>Treino: 350 artigos - 16.965 sentences (parágrafos)</li><li>Teste: 86 artigos – 3.525 sentences</li><li>Valid: 61 artigos – 2.235 sentences</li></ul>	<b>Task 2 (2019)</b> : SLC (single-label classif.) <ul style="list-style-type: none"><li>Classificar se <b>parágrafo</b> possui ou não propaganda(s) (<b>sim</b> ou <b>não</b>). <i>Melhor resultado: f1 = 0,67</i></li></ul> <b>Task TC (2021)</b> : MLC (multi-label classif.) <ul style="list-style-type: none"><li>Classificar as técnicas de persuasão contidas em um <b>fragmento de texto</b> (já identificado como propaganda)</li></ul>
2023	<b>Multilíngue – textos de artigos</b> <ul style="list-style-type: none"><li>23 classes</li><li>Treino/Teste/Valid: inglês, francês, alemão, italiano, polonês, russo</li><li>Apenas Teste: es, ka, gr</li></ul>	<b>Task 3</b> : MLC (multi-label classif.) <ul style="list-style-type: none"><li>Identificar as técnicas de persuasão presentes em cada <b>parágrafo</b>.</li></ul>
2024	<b>Multilíngue – textos de memes</b> <ul style="list-style-type: none"><li>20 classes</li><li>Treino/Teste/Valid: Inglês</li><li>Teste: Inglês + 3 ling. surpresa.</li><li><b>Hierarquia disponibilizada</b></li></ul>	<b>Task 1</b> : MLC (multi-label classif.) <ul style="list-style-type: none"><li>Identificar as técnicas de persuasão presentes em cada <b>texto de meme</b>.</li></ul>

### 1. Estatísticas Básicas da Base de Treino de 2024

#### 1.1 Estatísticas Gerais

	valor
N	7000
q	20
LCard	1,61285714
LDens	0,08064286
NDist	812
NUniq	452
PUniq	0,55665025
NMax	1264
max_labelsets	[()]
label_freqs	

'Appeal to authority': 850, 'Appeal to fear/prejudice': 337, 'Bandwagon': 97, 'Black-and-white Fallacy/Dictatorship': 780, 'Causal Oversimplification': 240, 'Doubt': 350, 'Exaggeration/Minimisation': 356, 'Flag-waving': 571, 'Glittering generalities (Virtue)': 488, 'Loaded Language': 1750, "Misrepresentation of Someone's Position (Straw Man)": 62, 'Name calling/Labeling': 1518, 'Obfuscation, Intentional vagueness, Confusion': 21, 'Presenting Irrelevant Data (Red Herring)': 59, 'Reductio ad hitlerum': 63, 'Repetition': 305, 'Slogans': 667, 'Smears': 1990, 'Thought-terminating cliché': 528, 'Whataboutism': 258
--

## 1.2 Labels x Tamanho do Texto

	media_caracteres	media_tokens	media_frases
geral	116,10	17,93	3,70
sem labels (não é propaganda)	57,41	8,82	2,79
Appeal to authority	194,92	30,50	4,30
Appeal to fear/prejudice	164,60	25,49	4,47
Bandwagon	92,24	14,33	3,65
Black-and-white Fallacy/Dictatorship	134,72	20,83	4,21
Causal Oversimplification	162,14	25,01	4,35
Doubt	166,59	25,92	4,33
Exaggeration/Minimisation	149,44	23,49	3,99
Flag-waving	128,22	20,24	3,94
Glittering generalities (Virtue)	158,80	24,75	4,18
Loaded Language	158,03	24,26	4,34
Misrepresentation of Someone's Position (Straw Man)	151,85	23,63	5,06
Name calling/Labeling	139,81	21,31	4,20
Obfuscation, Intentional vagueness, Confusion	190,52	31,52	4,52
Presenting Irrelevant Data (Red Herring)	157,14	24,53	4,44
Reductio ad hitlerum	171,59	24,10	5,17
Repetition	219,75	33,09	6,93
Slogans	116,10	17,63	4,13
Smears	148,14	22,81	4,27
Thought-terminating cliché	169,48	26,02	4,88
Whataboutism	184,28	27,97	4,78

### 1.3 Correlação entre os labels

Maiores valores:

- 0,13 (Glittering generalities (Virtue) x Flag-waving)
- 0,12 (Loaded Language x Name Calling)
- -0,11 (Doubt x Smears)

### 1.4 Resultado do meu Primeiro Teste (feito com a base de validação)

TF-IDF, stemming, sem stoplist, BR

f1-micro = 0,1538

### 1.5 Hierarquia

