



Universidade do Minho
Escola de Engenharia

LEI

Deteção de artigos de jornais e Posts que
despoletem controvérsia



Projeto 95

Constança Elias - PG42820

Jorge Brandão Gonçalves - PG42838

Maria Barbosa - PG42844

Pedro Pinheiro - PG44421

Coordenadores:

Professor Pedro Rangel Henriques

Professora Cristiana Araújo

Contextualização

Problema

- Importância de perceber se os **textos de artigos ou posts** dos *media*, em português, geram **controvérsia**, i.e., comentários negativos.
- Importância da contribuição deste estudo para a análise que existe nesta área em português.

Objetivo

- Analisar esses textos para avaliar a sua potencialidade para despoletar controvérsia;
- Procurar estabelecer uma relação entre a existência de **discurso pejorativo/grosseiro** nos comentários e o sentimento gerado pelo *post* (negativo, positivo ou neutro),

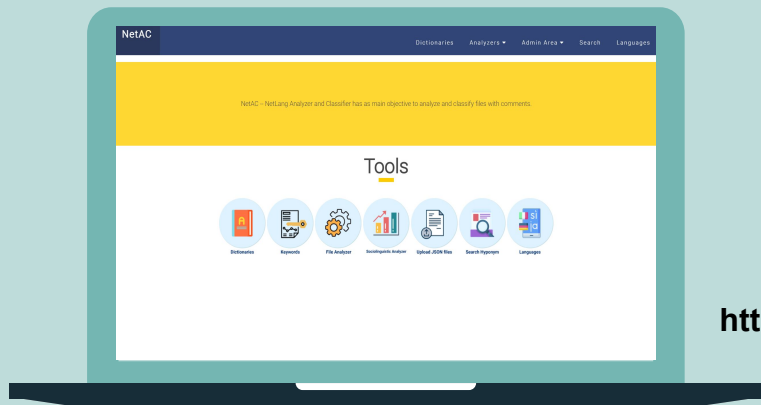
Proposta de Trabalho

- Utilizar como ponto de partida os **resultados** produzidos pela **ferramenta NetAc** (*NetLang Analyzer and Classifier*).
- Ter por base o **dataset** para análise de sentimentos do *Twitter*.
- Utilizar **modelos de ML**, para criar uma ferramenta capaz de identificar, a partir do conteúdo de um post, o tipo de reações que poderá provocar aos leitores.



Etapas Concluídas

1. Selecionar os jornais e notícias a avaliar (*Público* e *Sol*);
2. Preparar os ficheiros JSON com as notícias e comentários;
3. Submeter os JSON à ferramenta NetAC e tirar conclusões;



<http://netlang-corpus.ilch.uminho.pt:10100/>

Síntese dos Resultados obtidos

Jornal
Público

Table 1: Síntese dos resultados por ficheiro

Nome	N de Comentários	N de ocorrências de hate Speech	Percentagem de hate Speech(%)
Publico-extraction-portuguese-1.pdf	10	1	0,3067
Publico-extraction-portuguese-3.pdf	9	5	1,1682
Publico-extraction-portuguese-4.pdf	85	4	0,1044
Publico-extraction-portuguese-5.pdf	53	33	0,981
Publico-extraction-portuguese-6.pdf	9	0	0,0
Publico-extraction-portuguese-7.pdf	9	2	1,105
Publico-extraction-portuguese-8.pdf	12	7	1,3129
Publico-extraction-portuguese-9.pdf	88	25	0,5397
Publico-extraction-portuguese-10.pdf	15	3	0,3979
Publico-extraction-portuguese-11.pdf	8	1	0,2053
Publico-extraction-portuguese-12.pdf	13	0	0,0
Publico-extraction-portuguese-13.pdf	16	1	0,1451
Publico-extraction-portuguese-14.pdf	42	22	1,735
Publico-extraction-portuguese-15.pdf	43	5	0,3701
Publico-extraction-portuguese-16.pdf	31	14	0,6524
Publico-extraction-portuguese-17.pdf	6	3	1,3636
Publico-extraction-portuguese-18.pdf	15	5	0,6547
Publico-extraction-portuguese-19.pdf	5	2	1,6949
Publico-extraction-portuguese-20.pdf	9	3	1,6667
Publico-extraction-portuguese-21.pdf	7	2	1,4388
Publico-extraction-portuguese-22.pdf	3	0	0,0
Publico-extraction-portuguese-23.pdf	6	0	0,0
Publico-extraction-portuguese-24.pdf	6	0	0,0
Publico-extraction-portuguese-25.pdf	52	14	0,458
Publico-extraction-portuguese-26.pdf	18	2	0,4695

Síntese dos Resultados obtidos

Exemplo

“Um casal foi vítima de agressão **homofóbica**, com violência física a um deles, por parte de quatro homens, no domingo, pelas 19h40, no Terreiro do Paço, em Lisboa, disse esta segunda-feira a Polícia de Segurança Pública (PSP).”

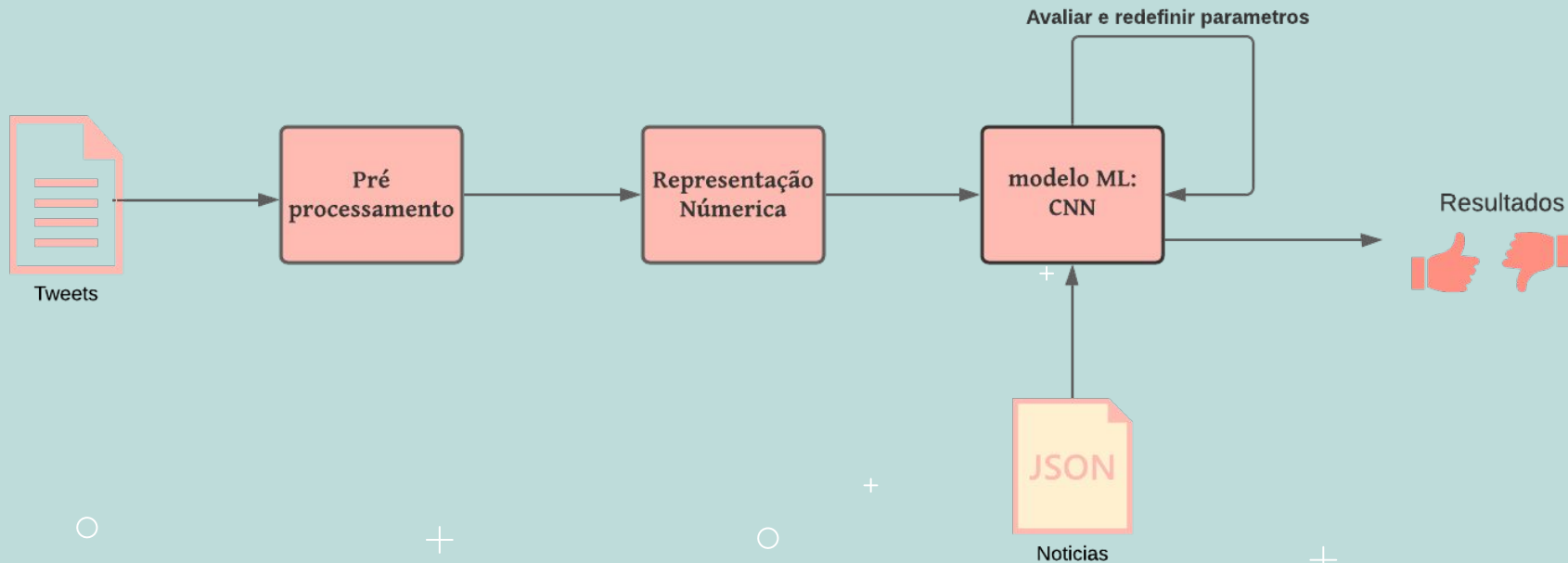
In www.publico.pt (2019/09/30)

Result analysis:

- Taking into account the words that were detected, we can reach the conclusion these comments are associated with : : Ethnicity - Black; Ethnicity - White; Sexual Identity - General; Ethnicity - General; Nationality - German; Ethnicity - Roma; Nationality - Chinese; Nationality - Japanese; Physical Identity - Physical (and Mental) Impairments; Gender - General;
- The percentage of hate speech related words is 0.981.
- Considering that the variable **Sexual Identity** General has the most occurrences in the post, we can interpret that this is the predominant hate speech.

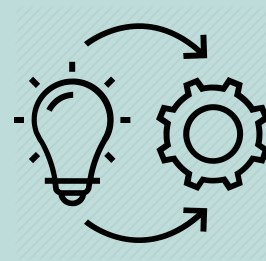
Etapas a Desenvolver

4. Criar um **modelo de ML** para avaliar se um Post gera sentimentos POSITIVOS, NEGATIVOS ou NEUTROS.



Etapas a Desenvolver

5. Criar uma **ferramenta** que faça uso do modelo de ML e que **retorne o tipo de sentimento gerado** pelo conteúdo da notícia submetida.
6. Desenvolver uma **aplicação** apelativa que **integre a ferramenta** desenvolvida.



Planeamento

Abril

- Submeter as notícias à ferramenta NetAC e extrair resultados;
- Pré-processar os dados e aplicar técnicas de PLN.
- Construir e analisar o modelo;
- Pesquisar quais os melhores hiperparâmetros do modelo;

Junho

- Desenvolver uma aplicação que integre a ferramenta desenvolvida.

Março

- Selecionar os jornais e as notícias;
- Preparar os JSON com notícias e comentários;
- Explorar as melhores estratégias a adotar para implementar o modelo de ML;
- Iniciar o tratamento do dataset;

Maio

- Melhorar o modelo;
- Transformar os json das notícias num formato .csv que possa ser aplicado no modelo;
- Criar uma ferramenta baseada no modelo de ML que recebe uma notícia e avalia o sentimento gerado.