



Universidade do Minho
Escola de Engenharia



Uma ferramenta para detecção de Notícias que
despoletam Controvérsia

2º Checkpoint de LEI 4/6/2021

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. Obter o *dataset* para a análise de sentimentos a partir do *Twitter* e classificá-lo. ✓
4. Definir e avaliar os resultados dos modelos de ML definidos. ✓
5. Desenvolver a aplicação Web que suporta a ferramenta de previsão da controvérsia de uma notícia. ✓
6. Incluir os temas mais “quentes” do momento. ✓
7. Começar a otimização de classificação do *dataset*, de forma a mantê-lo atualizado. ✓

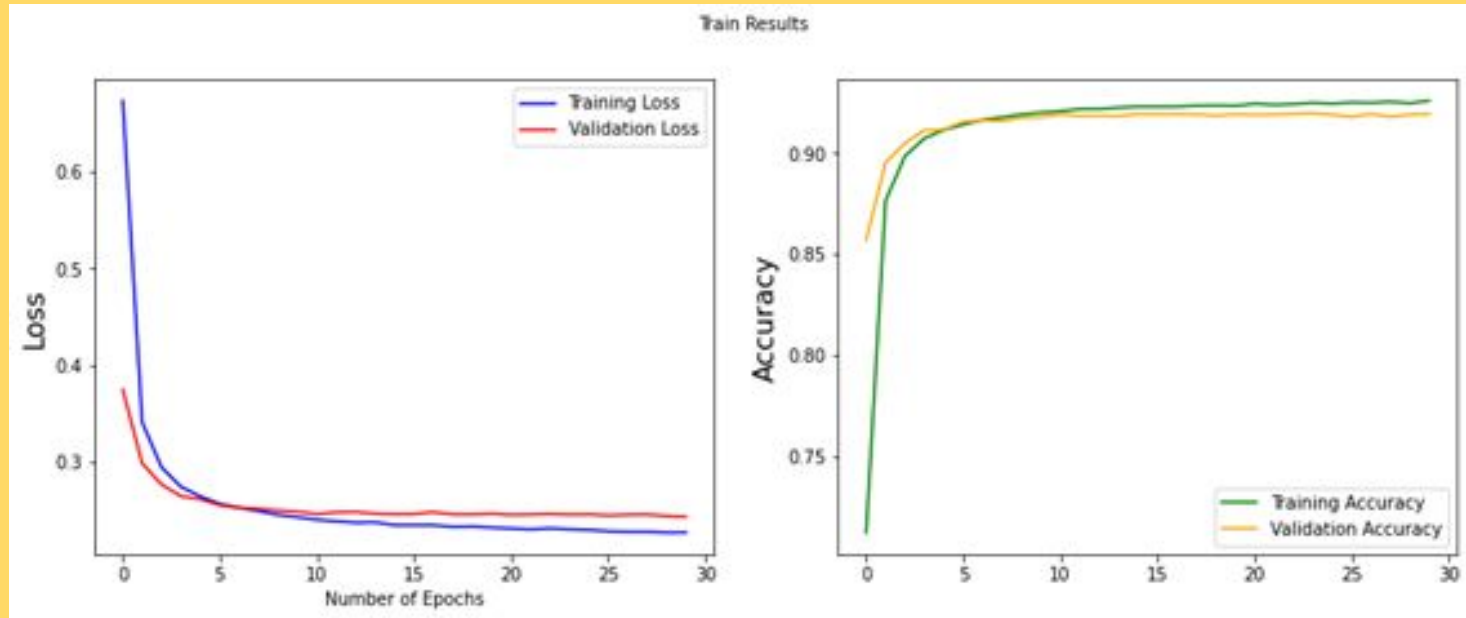
Resultados dos Modelos

- Valores de acurácia obtidos para os 4 principais modelos utilizados.

Modelo	accuracy
GRU	92.6%
LSTM	91%
Naive Bayes Model	78%
Logistic Regression Model	91%

Resultados da GRU

- Valores de *loss* e *accuracy* obtidos para o modelo GRU.



Validação de Resultados

- Com o *Corpus* de Notícias do **NetLang**.

Exemplo de alguns resultados:

1. ÁRABE PORCO E COVARDE BATENDO NA NOIVA - **Reativo**
2. BALANÇO GERAL - Racismo e prisão! PM negro é chamado de macaco - **Reativo**
3. Férias de luxo de cristina ferreira no dubai - **Reativo**

Demo

Planeamento

16/junho

- Validação dos resultados obtidos, usando o NetAc.
- Acabar o processo de atualização automática do modelo com base na classificação otimizada do *dataset*.

9/junho

- Finalizar a integração da ferramenta no servidor.
- Acabar a nova fase de classificação do dataset;

20/junho

- Entrega do Relatório