



Classificação automática de artigos estigmatizantes de doenças mentais em jornais de notícias portugueses *online* - Prémio Arquivo.pt 2022

Alina Yanchuk - Mestrado em Engenharia Informática

Orientadores: Doutora Alina Trifan
Doutor José Luís Oliveira
Doutora Olga Oliveira

Departamento de Eletrónica, Telecomunicações e Informática - Universidade de Aveiro

27 de Maio de 2022



Contexto e Motivação

A presença de estigma na nossa sociedade é ainda frequente — quando associado às doenças mentais, tem implicações negativas nos doentes, nos seus tratamentos e nos próprios profissionais de saúde.

A utilização de expressões estigmatizantes em textos publicados nos jornais de notícias é bastante comum — em muitos casos através de metáforas, ironia e sarcasmo.

Moção de Censura

Passos Coelho: país não precisa de viver todas as semanas na esquizofrenia da crise política

10.02.2011 - 18:09 Por Lusa

Votar ★★★★★ | 0 votos ★★★★★ 17 de 27 notícias em Política « anterior seguinte »

O líder do PSD, Pedro Passos Coelho, recusou hoje comentar o anúncio de uma moção ao Governo por parte do BE mas sublinhou que Portugal não pode viver "esquizofrenicamente todas as semanas" sentimentos de crise política.

Figura 1 — Exemplo de notícia com conteúdo estigmatizante.

Saúde

Esquizofrénicos passam a pagar uma parte dos medicamentos antipsicóticos

28.09.2010 - 07:36 Por Alexandra Campos

Votar ★★★★★ | 4 votos ★★★★★ 2 de 11 notícias em Sociedade « anterior seguinte »

Psiquiatras temem que mudança nas regras de comparticipação provoque diminuição da adesão à terapêutica e que doentes tenham recaídas.

Figura 2 - Exemplo de notícia com conteúdo não estigmatizante/literal.



Contexto e Motivação

A análise de artigos jornalísticos tem apresentado um grande crescimento na área da investigação, bem como a utilização de Inteligência Artificial para a realizar.

Fonte de dados?



O repositório Arquivo.pt permite aceder a um grande volume de notícias, recentes e antigas, e é também responsável pela realização do concurso **Prémio Arquivo.pt**.



Figura 3 — Prémio Arquivo.pt 2022.



Objetivos

Classificação automática do sentido de artigos de jornais de notícias *online*, recolhidos do **Arquivo.pt** e detentores de referências a **doenças mentais**, como **estigmatizante** ou **literal**.

Deteção automática de tópicos, através de *topic modeling*.



Doenças mentais e estigma



Figura 4 — Segundo a Ordem dos Psicólogos Portugueses, 1 em 5 portugueses sofrem de doenças mentais (2021).

A **Organização Mundial de Saúde** define o **estigma** como uma marca distintiva que, ao aliar-se às perturbações mentais, cria um ambiente de exclusão social e discriminação perante a pessoa estigmatizada.

Vários estudos (metodologia manual) **confirmam que os jornalistas contribuem para o desenvolvimento do estigma relativo à saúde mental.**

Espanha

695 notícias

47.9% notícias estigmatizantes

Reino Unido

600 notícias

11% notícias estigmatizantes

Brasil

229 notícias

34% notícias estigmatizantes

Estados Unidos da América

1740 notícias

28% notícias estigmatizantes



Os dados



Transtornos mentais em estudo:

ESQUIZOFRENIA

PSICOSE

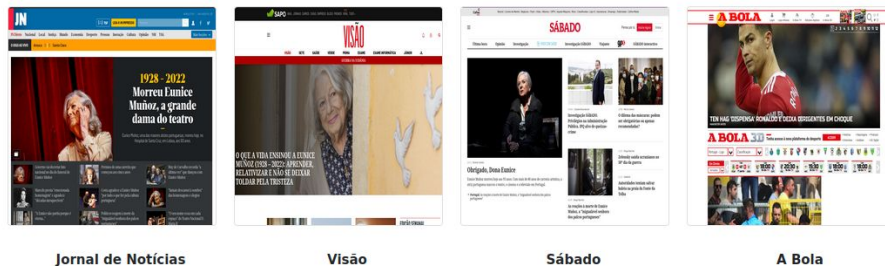
Jornais de notícias *online* utilizados:

Figura 5 – Jornais utilizados



Método de recolha:

Arquivo.pt API
+
Web scraping





Anotação manual

Todos os artigos recolhidos e relevantes para o problema foram classificados manualmente, por um conjunto de diferentes anotadores humanos, com as classes:

- **Estigmatizante**: o artigo utiliza a doença no sentido metafórico e dentro de um contexto inadequado, para revelar uma ideia que vai além do sentido literal do termo;
- **Literal**: o artigo utiliza a doença no seu sentido literal e dentro de um contexto adequado;



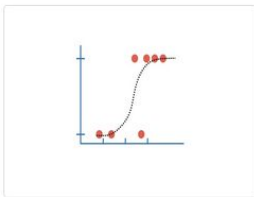
978 artigos manualmente classificados



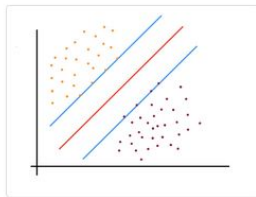
Modelos de classificação automática do sentido dos artigos

Modelos de representação: *Bag-of words* / TF-IDF / *Word embeddings* (GloVe 300D) / *Brazilian Portuguese LIWC 2007 Dictionary*

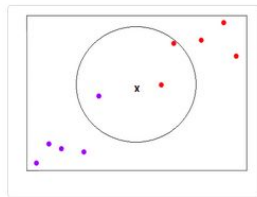
Machine learning:



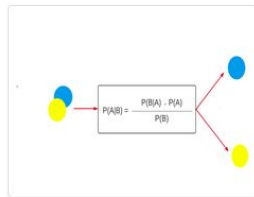
Logistic Regression (LR)



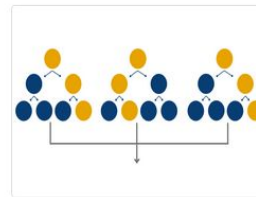
Support Vector Machine (SVM)



K-Nearest Neighbors (KNN)

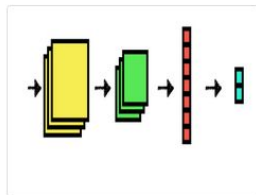


Naive Bayes (NB)

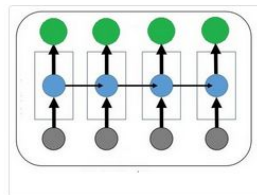


Random Forest (RF)

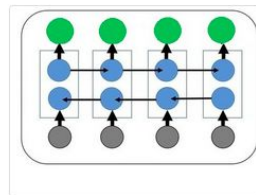
**Deep learning
(redes neurais):**



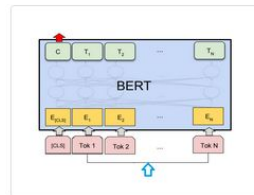
Convolutional Neural Network (CNN)



Long Short-Term Memory (LSTM)



Bidirectional LSTM (Bi-LSTM)



BERTimbau

Figura 6 -
Algoritmos
implementados



Detecção automática de tópicos

A detecção automática de tópicos foi realizada através de *topic modeling* usando o algoritmo **top2vec**, que foi treinado nos 978 documentos e permitiu obter:

- um conjunto de 50 termos descritivos dos tópicos descobertos e pontuações da sua similaridade ao tópico;
- os documentos semanticamente mais similares a cada tópico.



RESULTADOS - Classificação manual do sentido dos artigos

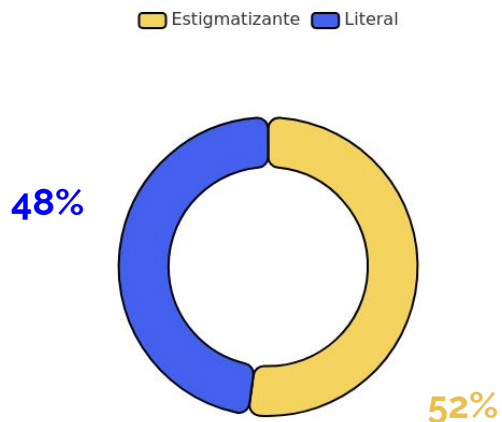


Figura 7 — Quantidade de artigos estigmatizantes e literais no total.

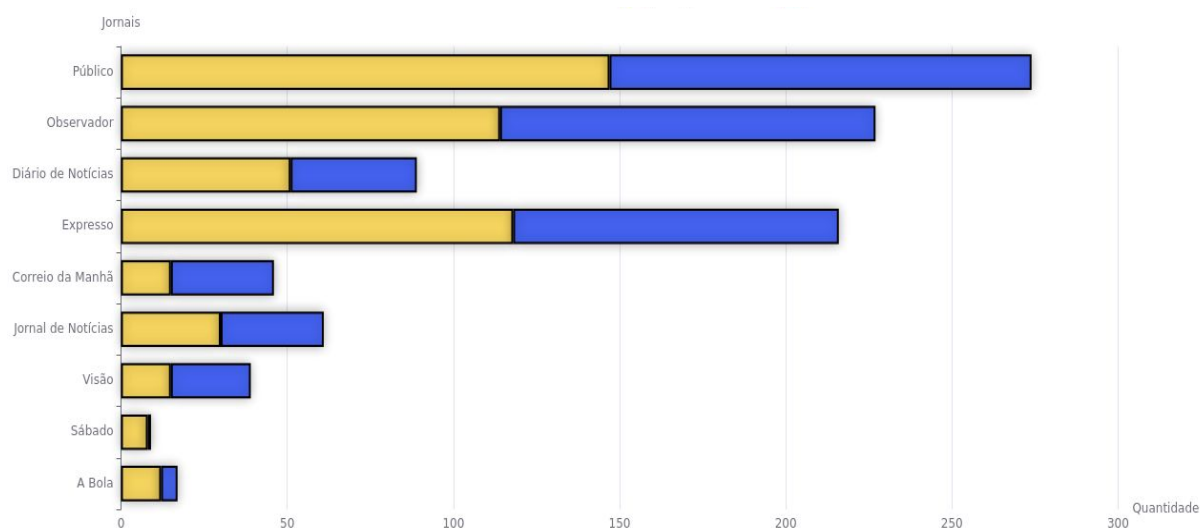


Figura 8 — Agrupamento por jornal de notícias.



RESULTADOS - Classificação automática do sentido dos artigos

Modelo de classificação	Modelo de representação	Exatidão (%)	Precisão	Recall	F1
<i>Logistic Regression</i>	<i>Bag-of-words</i>	92.35	0.92	0.93	0.93
	TF-IDF	93.37	0.93	0.94	0.94
	LIWC	70.41	0.73	0.69	0.71
SVM	<i>Bag-of-words</i>	90.31	0.92	0.90	0.91
	TF-IDF	90.82	0.93	0.90	0.91
	LIWC	80.10	0.81	0.80	0.81
<i>Naive Bayes</i>	<i>Bag-of-words</i>	91.33	0.91	0.92	0.92
	TF-IDF	93.37	0.91	0.97	0.94
	LIWC	52.04	0.52	1.00	0.69
KNN	<i>Bag-of-words</i>	65.82	0.89	0.40	0.54
	TF-IDF	91.33	0.92	0.91	0.92
	LIWC	70.41	0.72	0.70	0.71
<i>Random Forest</i>	<i>Bag-of-words</i>	92.86	0.90	0.97	0.93
	TF-IDF	91.84	0.88	0.97	0.93
	LIWC	79.08	0.77	0.86	0.81
CNN	<i>Word embeddings</i>	87.76	0.92	0.83	0.88
LSTM	<i>Word embeddings</i>	87.24	0.96	0.78	0.87
Bi-LSTM	<i>Word embeddings</i>	91.33	0.90	0.94	0.92
BERTimbau	<i>Tokenizer do BERTimbau</i>	91.84	0.93	0.91	0.92

Tabela 1 — Avaliação dos modelos implementados.



RESULTADOS - Detecção automática de tópicos

Foram automaticamente detetados **dez** tópicos — **as doenças mentais são, essencialmente, retratadas nas temáticas da saúde mental e quando associadas a ações criminais, e a maior percentagem de artigos estigmatizantes, relativamente ao total de artigos nesse tópico, está presente nos tópicos da Economia (97%) e da Política (96%).**

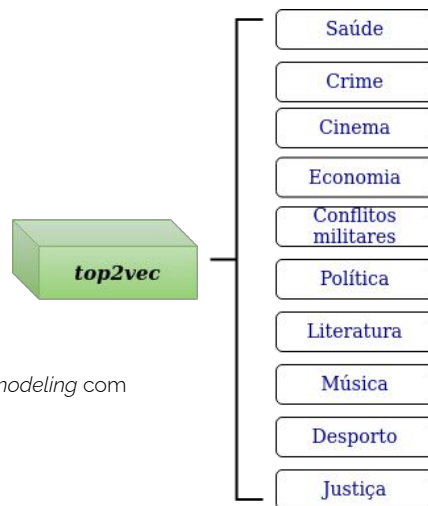


Figura 9 - Tópicos obtidos através de *topic modeling* com o algoritmo *top2vec*.



<https://alina-yanchuk02.github.io/estigma>



CLASSIFICARIA AUTOMÁTICA DE ARTIGOS ESTIGMATIZANTES DE DOENÇAS MENTAIS ONLINE

Expresso

ESTIGMATIZANTE LITERAL

DIÁRIO SE

EXPRESSO TRIBUNA BLUTZ VIDA EXTRA BOLA CAÍDA OS NEGRAS EXPRESSO MOBILIÁRIO O MARINHEIRO

- INÍCIO
- SOBRE O PROJETO
- EXEMPLO
- RESULTADOS
- TODOS OS ARTIGOS
- RECURSOS E CONTACTOS

Descoberta científica abre a porta à prevenção da esquizofrenia

19.01.2020 às 16h41

Investigador português em Londres comprovou que o cérebro destes doentes tem menos sinapses do que o das pessoas saudáveis. A investigação tenta agora restaurar as ligações entre os neurónios, abrindo caminho à prevenção e tratamento de uma das mais graves e incapacitantes doenças mentais

CHRISTIANA MARTINS

O psiquiatra Tiago Reis Marques, de 43 anos, integra a equipa que esta semana publicou a descoberta D.R.

CLASSIFICARIA AUTOMÁTICA DE ARTIGOS ESTIGMATIZANTES DE DOENÇAS MENTAIS ONLINE

RECICLO
SOBRE O PROJETO
EXEMPLO
RESULTADOS
TODOS OS ARTIGOS
RECURSOS E CONTACTOS

universidade de aveiro

FCT Fundação para a Ciência e a Tecnologia

LER ARTIGO COMPLETO

Artigos por Tópico

TÓPICO 0 TÓPICO 1 TÓPICO 2 TÓPICO 3 TÓPICO 4 TÓPICO 5 TÓPICO 6 TÓPICO 7 TÓPICO 8 TÓPICO 9

Saúde 232 artigos

Tendo em conta os 50 termos mais descritivos, detetados automaticamente, os artigos pertencentes a este tópico parecem abordar o tema da Medicina e Saúde:

Tip 3 artigos com sentido **estigmatizante**

#	Título	Acesso
1	Rocher pronta a obter produção de Tamiflu à concorrência	Aceder
2	Españoles quieren pór fin a "esquizofrenia que reina na costa"	Aceder
3	Pais tamen ramos des de cancro en portada de escola en Braga	Aceder

Tip 3 artigos com sentido **literal**

#	Título	Acesso
1	Consumo de cannabis e esquizofrenia têm una forte relación genética	Aceder
2	Esquizofrenia pode resultar de alteração num gene	Aceder
3	Inferimentos por surto psicótico em consumidores de cannabis aumentou	Aceder

Figura 10 - Exemplos de páginas do *website*.



Conclusão e Contribuições do projeto

- Contribui para a **sensibilização para o problema da presença de estigma** na imprensa portuguesa;
- Disponibiliza um conjunto de **978 artigos anotados** e **9 algoritmos de Inteligência Artificial** para a tarefa de classificação;
- **Apresenta um nível inovador**, ao explorar a estigmatização de doenças mentais na comunicação social portuguesa através de Inteligência Artificial;
- **Tema e metodologia pioneiros** no âmbito do concurso Prémio Arquivo.pt;
- Os modelos desenvolvidos neste projeto podem ser estendidos e ajustados a outros problemas similares — **referência para futuros trabalhos, sobretudo na área de Processamento de Linguagem Natural em português**.



Referências

Arquivo.pt: <https://arquivo.pt/>

Artigo científico/técnico e dados anotados: <https://drive.google.com/drive/folders/1HwqpYzgixXrQAeMVGnh5lpfLiiFtTV3->

Projeto de código-aberto: <https://github.com/alina-yanchuk02/stigmaClassification>

OBRIGADA!