



RICARDO SAMPAIO SOUSA JUNIOR ALENCAR

MODELO DE TCC - UNICHRISTUS

FORTALEZA

2023

RICARDO SAMPAIO SOUSA JUNIOR ALENCAR

MODELO DE TCC - UNICHRISTUS

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Sistemas de Informação do Centro Universitário Christus, como requisito parcial para obtenção do grau de bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Daaaaaaaaaaaaaaaa-
aaa

FORTALEZA

2023

RICARDO SAMPAIO SOUSA JUNIOR ALENCAR

MODELO DE TCC - UNICHRISTUS

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Sistemas de Informação do Centro Universitário Christus, como requisito parcial para obtenção do grau de bacharel em Sistemas de Informação.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Daaaaaaaaaaaaaaaaaaaa (Orientador)
Centro Universitário Christus (Unichristus)

Prof. Ms. Daaaaaaaaaaaaaaaaaaaa
Centro Universitário Christus (Unichristus)

Prof. Ms. Feeeeeeeeeeeeeeeeee
Centro Universitário Christus (Unichristus)

AGRADECIMENTOS

"Que nada nos defina, que nada nos sujeite. Que a liberdade seja a nossa própria substância."

(Simone de Beauvoir)

RESUMO

Com a popularização da internet e conseqüentemente com a facilidade

Palavras-chave: Fake News. Aprendizado de Máquina. Processamento de Linguagem Natural.

ABSTRACT

With the internet popularization and consequently

Keywords: Fake News. Machine Learning. Natural Language Processing.

LISTA DE FIGURAS

Figura 1 – Exemplo de Falsa Conexão.	14
--	----

LISTA DE TABELAS

Tabela 1 – Conjunto de dados no formato atributo-valor.	15
---	----

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Contextualização e delimitação do tema	11
1.2	Problematização	11
1.3	Pressupostos	11
1.4	Objetivos	11
1.4.1	<i>Objetivo geral</i>	11
1.4.2	<i>Objetivos específicos</i>	11
1.5	Justificativa	12
1.6	Estrutura do trabalho	12
2	REFERENCIAL TEÓRICO	13
2.1	As fake news e sua disseminação por meio das mídias digitais	13
2.2	Tipologia das fake news	13
2.2.1	<i>Sátira ou paródia</i>	13
2.2.2	<i>Falsa conexão</i>	13
2.2.3	<i>Conteúdo enganoso</i>	13
2.2.4	<i>Falso contexto</i>	13
2.2.5	<i>TF-IDF</i>	13
2.2.6	<i>Scikit-Learn</i>	14
2.3	Aprendizado de máquina	15
2.3.1	<i>A hierarquia do aprendizado</i>	15
2.3.2	<i>Aprendizado de máquina supervisionado</i>	15
2.4	Aprendizado de máquina supervisionado e classificadores	15
2.4.1	<i>Naive Bayes</i>	15
2.5	Algoritmos de fragmentação	15
2.5.1	<i>Algoritmo Hold-Out</i>	15
2.5.2	<i>Algoritmo Cross Validation K-Fold</i>	15
3	TRABALHOS CORRELATOS	17
4	METODOLOGIA	18
5	RESULTADOS	19
6	CONCLUSÕES E TRABALHOS FUTUROS	20

REFERÊNCIAS 21

1 INTRODUÇÃO

1.1 Contextualização e delimitação do tema

Segundo a definição do dicionário Aurélio (FERREIRA, 2014) entende-se por notícia o relato de acontecimento atual, de interesse público geral, ou de determinado segmento da sociedade, veiculado em jornal, rádio, televisão, etc. Essas notícias geralmente são repassadas por fontes oficiais, que obedecem um conjunto de normas éticas com o intuito de garantir que as informações sejam divulgadas sem alterações ou detalhes falsos. Hoje, com a popularização da internet, o número de informações produzidas cresce em grandes proporções. De acordo com um estudo feito por Barbosa (2012), a cada dia do ano de 2012 o mundo produziu mais informação escrita do que toda a que existia antes de 2003.

Essas contas são provavelmente *bots*, e descobrimos várias estratégias de manipulação que usam. Primeiro, os *bots* são particularmente ativos em amplificar notícias falsas no estágio inicial de propagação, antes que o conteúdo seja viral. Em segundo, os *bots* miram usuários influentes através de respostas e menções. (SHAO *et al.*, 2017).

1.2 Problematização

Desse modo tem-se como pergunta de pesquisa "Como elaborar uma aplicação que seja capaz de auxiliar no trabalho investigativo jornalístico e também no combate a desinformação gerada pela disseminação de informações falsas na web?".

1.3 Pressupostos

1.4 Objetivos

1.4.1 *Objetivo geral*

1.4.2 *Objetivos específicos*

1. Utilizar *fake news*.
2. Produzir *Multinomial NB* e *Passive Aggressive*
3. utilizando as técnicas *Cross-Validation* e *Hold-Out*.
4. Desenvolver

1.5 Justificativa

1.6 Estrutura do trabalho

2 REFERENCIAL TEÓRICO

2.1 As *fake news* e sua disseminação por meio das mídias digitais

As *fake news* constituem uma espécie de "imprensa marrom" (*yellow press*) intencionalmente veiculando conteúdos falsos, sempre com a intensão de obter algum tipo de vantagem, seja financeira (mediante receitas oriundas de anúncios), política ou eleitoral (CARVALHO; KANFFER, 2018).

tentou manipular a eleição pontifícia de 1522, escrevendo sonetos perversos sobre todos os candidatos (exceto o seu favorito, Medici) e colando-os para o público admirar no busto de uma figura conhecida como Pasquino, em Roma. O "pasquinade" então se transformou em um gênero comum de difundir notícias desagradáveis, a maioria delas falsas, sobre figuras públicas.

2.2 Tipologia das *fake news*

Com a popularização das mídias sociais nos últimos anos e, conseqüentemente, com o aumento da criação de notícias falsas, foi possível notar que as *fake news* possuem propriedades e atributos diferentes. Wardle (2017) definiu essa variedade de notícias em sete categorias diferentes como forma de classificar e facilitar a detecção das mesmas, são elas:

1. Sátira ou paródia
2. Falsa conexão

2.2.1 *Sátira ou paródia*

2.2.2 *Falsa conexão*

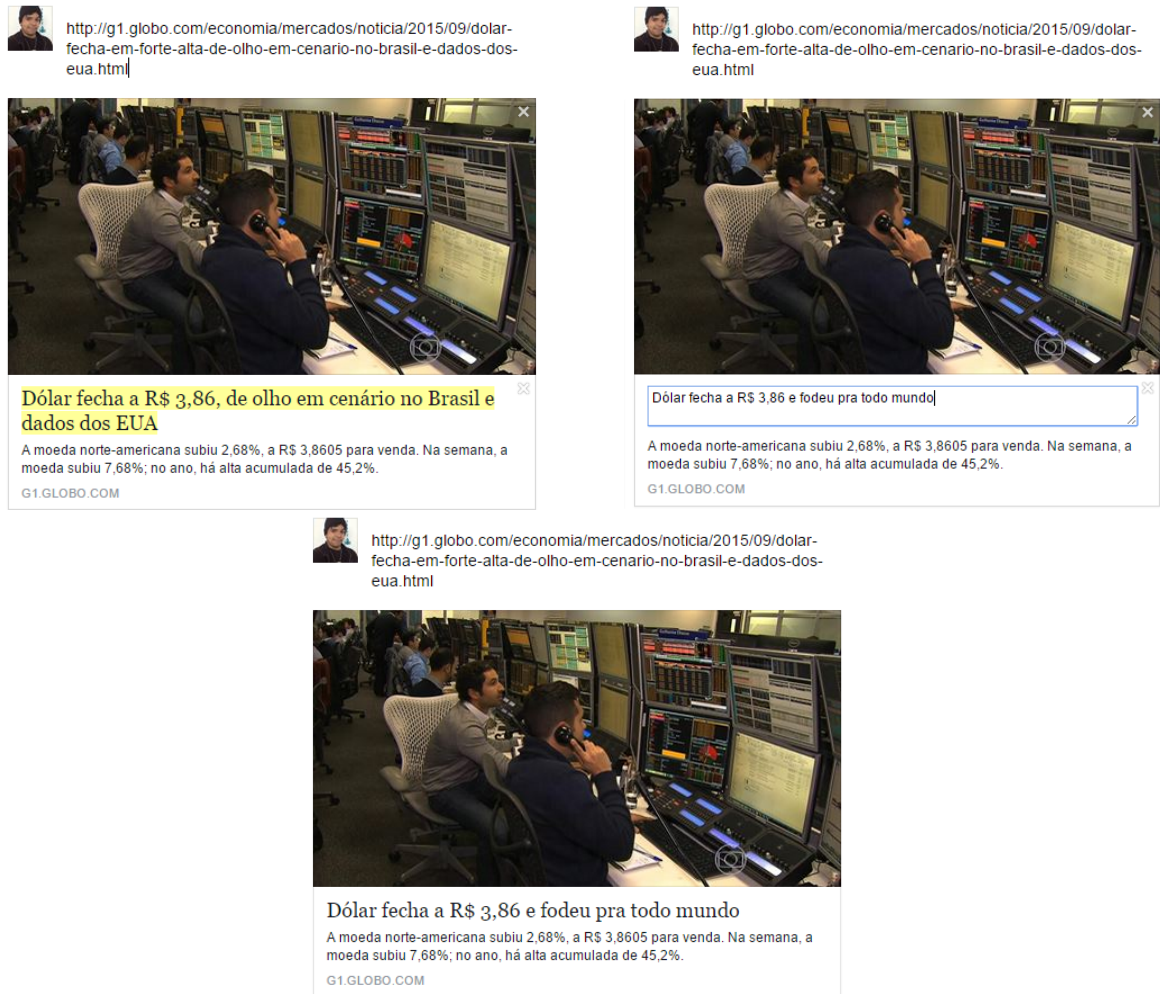
2.2.3 *Conteúdo enganoso*

2.2.4 *Falso contexto*

2.2.5 *TF-IDF*

O cálculo do TF-IDF dá-se a partir da Equação 2.1, sendo t_j a frequência de um termo em dado documento d_i multiplicado pelo fator de ponderação, que varia entre 0 e o $\log N$, onde N é o número de documentos do conjunto de dados e $d(t_j)$ é o número de documentos onde

Figura 1 – Exemplo de Falsa Conexão.



Fonte: Camarossi (2015).

há a ocorrência do termo pelo menos uma vez (CAVALCANTI *et al.*, 2011).

$$tfidf(t_j, d_i) = freq(t_j, d_i) \times \log \frac{N}{d(t_j)} \quad (2.1)$$

2.2.6 Scikit-Learn

Por se tratar de uma biblioteca para mineração e análise de dados, a mesma conta com uma série de pacotes necessários para o seu aproveitamento, como por exemplo, o `matplotlib`¹, o `numpy`² e o `pandas`³. Neste projeto, todos os classificadores utilizados e testados foram providos pelo `scikit-learn`.

¹ <<https://matplotlib.org/>>

² <<https://www.numpy.org/>>

³ <<https://pandas.pydata.org/>>

2.3 Aprendizado de máquina

2.3.1 A hierarquia do aprendizado

2.3.2 Aprendizado de máquina supervisionado

No aprendizado supervisionado é fornecido a um algoritmo de aprendizado, denominado indutor, um conjunto de dados de treinamento $E = \{E_1, E_2, \dots, E_N\}$, onde cada exemplo $E_i \in E$ possui um rótulo associado que define a classe a qual o mesmo pertence. Formalmente, pode-se dizer que cada exemplo $E_i \in E$ pode ser representado por uma tupla

$$E_i = (\vec{x}_i, y_i) \quad (2.2)$$

Tabela 1 – Conjunto de dados no formato atributo-valor.

	A_1	A_2	...	A_M
E_1	x_{11}	x_{12}	...	x_{1M}
E_2	x_{21}	x_{22}	...	x_{2M}
\vdots	\vdots	\vdots	\vdots	\vdots
E_N	x_{N1}	x_{N2}	...	x_{NM}

Fonte: Batista e Monard (2004).

2.4 Aprendizado de máquina supervisionado e classificadores

2.4.1 Naive Bayes

2.5 Algoritmos de fragmentação

2.5.1 Algoritmo Hold-Out

2.5.2 Algoritmo Cross Validation K-Fold

A partir da matriz de confusão pode-se extrair algumas medidas que podem ser utilizadas para avaliação de desempenho do modelo de classificação (FACELI *et al.*, 2011), são elas:

- **Acurácia:** Indica a per

$$ac(\hat{f}) = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.3)$$

3 TRABALHOS CORRELATOS

4 METODOLOGIA

5 RESULTADOS

6 CONCLUSÕES E TRABALHOS FUTUROS

REFERÊNCIAS

- BARBOSA, P. **Harvard trends–Tendências de gestão 2013**. [S.l.]: Porto: Vida Económica, 2012.
- BATISTA, G.; MONARD, M. C. Sniffer: um ambiente computacional para gerenciamento de experimentos de aprendizado de máquina supervisionado. **Proceedings of the I WorkComp Sul**, 2004.
- CAMAROSSO, F. **O dia em que descobri que ninguém clica na notícia**. 2015. Disponível em: <<https://medium.com/@fellipecamarossi/o-dia-em-que-descobri-que-ningu%C3%A9m-clica-na-not%C3%ADcia-eab77721e98f>>. Acesso em: 11 mar. 2020.
- CARVALHO, G. A. C. L. de; KANFFER, G. G. B. **O tratamento jurídico das notícias falsas (fake news)**. 2018. Disponível em: <<https://www.conjur.com.br/2018-mar-19/opinio-legislacao-dispoe-ferramentas-combater-fake-news>>. Acesso em: 09 mar. 2020.
- CAVALCANTI, E. R.; CAVALCANTI, E. P.; PIRES, C. E.; COSTA, R. A.; CAVALCANTI, C. R. Detecção e avaliação de cola em provas escolares utilizando mineração de texto: um estudo de caso. **Revista Brasileira de Informática na Educação**, v. 19, n. 02, p. 56, 2011.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. *et al.* Inteligência artificial: Uma abordagem de aprendizado de máquina. 2011.
- FERREIRA, A. B. d. H. Novo dicionário aurélio da língua portuguesa. In: POSITIVO (Ed.). **Novo dicionário Aurélio da língua portuguesa**. 5. ed. [S.l.: s.n.], 2014.
- SHAO, C.; CIAMPAGLIA, G. L.; VAROL, O.; FLAMMINI, A.; MENCZER, F. The spread of fake news by social bots. **CoRR**, abs/1707.07592, 2017. Disponível em: <<http://arxiv.org/abs/1707.07592>>.
- WARDLE, C. The different types of mis and disinformation. 2017. Disponível em: <<https://firstdraftnews.org/latest/fake-news-complicated/>>. Acesso em: 10 mar. 2020.