

## **VIOLÊNCIA DOMÉSTICA: UTILIZAÇÃO DE IA PARA QUANTIFICAÇÃO DAS OCORRÊNCIAS NOTICIADAS NA INTERNET**

### ***DOMESTIC VIOLENCE: USE OF AI TO QUANTIFY OCCURRENCES REPORTED ON THE INTERNET***

### ***VIOLENCIA DOMÉSTICA: USO DE LA IA PARA CUANTIFICAR LOS HECHOS DENUNCIADOS EN INTERNET***

**Edilene A. Veneruchi de Campos<sup>1</sup>**  
**Ana Cláudia De Oliveira Pedro Andréo<sup>2</sup>**

**RESUMO:** Mesmo com o aumento de denúncias sobre a violência contra a mulher, existe uma dificuldade muito grande de contabilizar estes delitos. Muitas vezes, não existe uma integração eficiente entre as autoridades, o que leva à necessidade de maneiras mais eficazes para quantificar estes números. Para minimizar este problema, este artigo trata da Análise de Sentimentos contidas em notícias publicadas em jornais de língua portuguesa. O objetivo é automatizar a busca de publicações contendo episódios de violência contra a mulher, visando gerar indicadores mais atualizados. Os dados utilizados relacionaram-se com o tema violência contra a mulher, pois estas podem ser utilizadas como indicadores para ações de enfrentamento deste problema, tendo sido utilizados dois sites de notícia. O primeiro passo foi a obtenção das notícias através do processo de *web crawling*, podendo-se filtrar as notícias relativas ao tema e criando-se duas nuvens de *tags*: notícias relacionadas à mulher e notícias associadas à violência contra a mulher, servindo como entrada para o treinamento de uma rede neural artificial. Durante os treinamentos realizados, viu-se como contribuição, em trabalhos futuros, a possibilidade de aumento de dados, sendo importante analisar a qualidade do *dataset* utilizado, tanto em propriedade das notícias quanto em quantidade.

**PALAVRAS-CHAVE:** Violência contra a Mulher; Análise de Sentimentos; Redes Neurais Artificiais.

**ABSTRACT:** Even with the increase in reports of violence against women, there is great difficulty in counting these crimes. Many times, there is no efficient integration among authorities, which leads to the need for more effective ways to quantify these numbers. To minimize this problem, this article deals with Sentiment Analysis

---

1 Edilene Aparecida Veneruchi de Campos é professora da Faculdade Insted. ORCID iD: <https://orcid.org/0000-0002-4427-608X>. E-mail: [edilene.veneruchi@insted.edu.br](mailto:edilene.veneruchi@insted.edu.br).

2 Ana Cláudia De Oliveira Pedro Andréo é professora do Instituto de Educação, Ciência e Tecnologia do Mato Grosso do Sul (IFMS). ORCID iD: <https://orcid.org/0009-0000-6593-4259>. E-mail: [ana.andreo@ifms.edu.br](mailto:ana.andreo@ifms.edu.br).

contained in news published in Portuguese-language newspapers. The objective is to automate the search for publications containing episodes of violence against women, aiming to generate more up-to-date indicators. The data used was related to the theme of violence against women, as these can be used as indicators for actions to combat this problem, using two news websites. The first step was to obtain the news through the web crawling process, being able to filter the news related to the topic and creating two tag clouds: news related to women and news associated with violence against women, serving as input for training an artificial neural network. During the training carried out, the possibility of increasing data was seen as a contribution in future work, and it is important to analyze the quality of the dataset used, both in terms of news property and quantity.

**KEYWORDS:** Violence against Women; Sentiment Analysis; Artificial Neural Networks.

**RESUMEN:** Incluso con el aumento de las denuncias de violencia contra las mujeres, es muy difícil dar cuenta de estos crímenes. De hecho, no existe una integración eficiente entre autoridades, lo que lleva a la necesidad de encontrar formas más efectivas de cuantificar estas cifras. Para minimizar este problema, este artículo aborda Análisis de Sentimiento contenido en noticias publicadas en periódicos en lengua portuguesa. El objetivo es automatizar la búsqueda de publicaciones que contengan episodios de violencia contra mujeres, con objetivo de generar indicadores más actualizados. Los datos utilizados están relacionados con el tema de la violencia contra mujeres, ya que pueden ser utilizados como indicadores de acciones para combatir este problema, utilizando sitios web de noticias. El primer paso fue obtener noticias a través del proceso de rastreo web, pudiendo filtrar noticias relacionadas con el tema y creando dos nubes de etiquetas: noticias relacionadas con mujeres y noticias asociadas a violencia contra mujeres, sirviendo como insumo para entrenamiento de la red neuronal artificial. Durante la formación realizada se vislumbra la posibilidad de incrementar datos como una aportación a trabajos futuros, analizando la calidad del conjunto de datos utilizados, tanto en términos de propiedad de las noticias como de cantidad.

**PALABRAS CLAVE:** Violencia contra la Mujer; Análisis de Sentimiento; Redes neuronales artificiales.

## INTRODUÇÃO

A violência contra a mulher no Brasil é um problema que se manifesta de diversas formas, como estupro, feminicídios e agressões. O país ocupa lugares no topo no ranking mundial de violência contra a mulher.

A Lei Maria da Penha, criada em 2006, foi uma medida para coibir a violência doméstica e facilitar a denúncia de agressões.

Segundo a Subsecretaria de Políticas Públicas para Mulheres (Violência, 2024):

Mato Grosso do Sul é um estado pioneiro na defesa dos direitos das mulheres. O primeiro organismo de políticas para mulheres foi criado em janeiro/1999, como "Coordenadoria Especial de Políticas para Mulheres" (Lei Estadual nº 4.928). Em abril/2003 foi transformado em "Subsecretaria de Políticas Públicas para Mulheres e Promoção da Igualdade Racial" (Lei Estadual nº 4.331) e desde 2015 existe como "Subsecretaria de Políticas Públicas para Mulheres".

A Aprendizagem de Máquina (*machine learning*) tem-se tornado popular em diversas áreas, como diagnósticos médicos; justiça e segurança pública; recrutamento e seleção de recursos humanos; comércio eletrônico; agricultura de precisão; dentre outros.

Em problemas relacionados a processamento de linguagem natural ou a reconhecimento de imagens, tem sido usado um subconjunto de machine learning, chamado Aprendizagem Profunda (*deep learning*). Diferentemente da aprendizagem de máquina convencional, aprendizagem profunda consegue detectar padrões complexos e abstratos, aprendendo representações e recursos úteis a partir de dados brutos.

O objetivo deste trabalho é investigar, analisar e testar modelos para análise e classificação de textos. Os dados utilizados relacionavam-se com o tema violência contra a mulher, visto que, a quantidade de notícias veiculadas pode ser utilizada como indicadores para ações de enfrentamento deste problema tão sério. As seções a seguir apresentam breve fundamentação teórica, os materiais e métodos empregados, análise de resultados e comentários conclusivos, apontando trabalhos futuros. O código desenvolvido para a realização das atividades aqui descritas, bem como os *dataset* utilizados encontram-se disponíveis em em uma pasta compartilhada.<sup>3</sup>

## DESENVOLVIMENTO

---

<sup>3</sup> Código desenvolvido: <https://drive.google.com/drive/folders/1Zql2-URe5reUwU7SULab-zxt7Yy4KY?usp=sharing>

Em fevereiro de 2024, a Agência Senado divulgou uma pesquisa de violência contra a mulher nos estados e no DF (Agência Senado, 2024), que afirma que:

[...]a percepção de que a violência doméstica aumentou nos últimos 12 meses é majoritária em todo o país (74%), com algumas variações nas unidades federativas. O maior percentual de mulheres que afirmam que a violência doméstica aumentou está no Distrito Federal (84%), e o menor, no Rio Grande do Sul (62%).

Infelizmente, a Agência Brasil corrobora estes dados (Campos, 2024):

No ano de 2023, ao menos oito mulheres foram vítimas de violência doméstica a cada 24 horas. Os dados referem-se a oito dos nove estados monitorados pela Rede de Observatórios da Segurança (BA, CE, MA, PA, PE, PI, RJ, SP). A informação consta do novo boletim *Elas Vivem: Liberdade de Ser e Viver*, divulgado nesta quinta-feira (7). Ao todo, foram registradas 3.181 mulheres vítimas de violência, representando um aumento de 22,04% em relação a 2022, quando Pará e Amazonas ainda não faziam parte deste monitoramento. Ameaças, agressões, torturas, ofensas, assédio, feminicídio. São inúmeras as violências sofridas que não começam ou se esgotam nas mortes registradas. Os dados monitorados apontaram 586 vítimas de feminicídios. Isso significa dizer que, a cada 15 horas, uma mulher morreu em razão do gênero, majoritariamente pelas mãos de parceiros ou ex-parceiros (72,7%), que usaram armas brancas (em 38,12% dos casos), ou por armas de fogo (23,75%).

Também de acordo com a Agência Senado, 2024,

Menos de um quarto das brasileiras (24%) afirma conhecer muito sobre a Lei Maria da Penha (Lei 11.340, de 2006). Esse índice é maior no Distrito Federal (33%), e menor no Piauí (17%), Maranhão (17%), Amazonas (17%), Paraíba (19%) e Pará (19%). A pesquisa mostra que cerca de metade das brasileiras acredita que a Lei Maria da Penha protege apenas em parte as mulheres contra a violência doméstica e familiar (51%), patamar semelhante ao das unidades federativas do país.

Pode-se notar que, apesar do aumento da conscientização, por parte da sociedade, da violência contra a mulher, ainda se necessita de uma forma mais abrangente para que os números de casos sejam detectados.

Muito da popularidade de aplicações relacionadas com aprendizado de máquina está relacionada com o uso de Redes Neurais Profundas (*Deep Neural Networks* - DNN) e com o paradigma de aprendizagem orientada a dados, que permitem aos modelos aprenderem representações úteis automaticamente a partir dos dados. Em se tratando de dados textuais, o Processamento de

Linguagem Natural - PLN trata de necessidades de extração, processamento e classificação de opiniões, emoções e sentimentos da imensa quantidade de conteúdo disponível na *Web*. Esta seção aborda, sucintamente, temas importantes para melhor delinear o trabalho desenvolvido.

## **1. Redes Neurais e Redes Neurais Profundas**

Rede Neural Artificial (RN) é um modelo computacional inspirado nas redes neurais humanas, formado por um conjunto de neurônios conectados em um grafo, que pode ser cíclico (recorrente) ou acíclico (*feedforward*). A organização dos neurônios, normalmente, se dá em camadas e, assim, um neurônio recebe entradas vindas de neurônios da camada anterior via dendritos (processo de sinapse), realiza o processamento necessário (função de ativação) e envia o resultado para neurônios da camada posterior, através do axônio. RN são formadas por camada de entrada, camada de saída e um conjunto de camadas intermediárias, chamadas camadas ocultas (Faceli, 2011).

Aprendizagem profunda está relacionada com redes neurais profundas (*Deep Neural Network* - DNN) que recebem este nome por possuírem muitas camadas ocultas. Uma DNN consegue obter vários níveis de abstração, melhorando o processamento de textos, imagens, vídeos e áudios (Goodfellow, 2016).

## **2. Processamento de Linguagem Natural**

O Processamento de Linguagem Natural - PLN - permite a automatização da realização de tarefas relacionadas com a linguagem. Assim, muitos trabalhos têm sido feitos para permitir que máquinas consigam ler, ouvir e interpretar mensagens, identificar sentimentos nelas contidos e definir graus de importância, por exemplo.

## **3. Análise de Sentimento**

A Análise de Sentimento é uma sub-área do Processamento de Linguagem Natural e visa à extração, o processamento e à classificação de opiniões, emoções, sentimentos expressos em conteúdos disponíveis na *Web*. A

Análise de Sentimento, por sua vez, se subdivide em três níveis: 1) nível de documento, quando um sentimento é atribuído ao documento inteiro; 2) nível de sentença, quando um sentimento é atribuído a cada sentença do documento; 3) nível de aspecto, quando um sentimento é atribuído a todos os aspectos de uma entidade (pessoa, objeto, produto, etc) identificada. (Liu, 2012).

#### **4. Transferência de Aprendizagem e Aumento de dados**

Transferência de aprendizagem é a reutilização do conhecimento já adquirido durante o treinamento de um modelo. Este processo é utilizado para reduzir o esforço necessário que se tem ao treinar um modelo do zero. Silva, 2018, cita que existem 3 formas da transferência de conhecimento ser realizada: extração de características (*feature extraction*); afinação (*fine tuning*); treinamento conjunto (*joint training*).

Aumento de dados é uma técnica utilizada para gerar novos exemplares de dados de treinamento a fim de aumentar a generalidade do modelo. Exemplos desta técnica são transformações geométricas, como rotação, translação, escala, corte, dentre outros, em caso de imagens e, para textos, algumas técnicas são *word embedding*, *back translation*, *thesaurus*, dentre outras.

#### **4. Conjunto de dados**

Considerando que o presente trabalho teve como propósito promover a aprendizagem e o entendimento da utilização de modelos para análise e classificação de texto, optou-se pela construção de um *dataset* próprio para entendimento de todas as necessidades embutidas nessa fase. Isto se mostrou extremamente importante, uma vez que a qualidade dos dados processados tem impacto direto nos resultados obtidos.

Utilizou como fonte de dados dois sites de notícia: o BBC *News* Brasil<sup>4</sup>,

---

<sup>4</sup> <https://www.bbc.com/news/portuguese>

com abrangência mundial, e o Campo Grande *News*<sup>5</sup>, cujas notícias abrangem o Estado de Mato Grosso do Sul.

O primeiro passo nesta fase foi a obtenção das notícias através do processo popularmente chamado de *web crawling*. Para o portal *BBC News* foram empregados recursos da biblioteca *BeautifulSoup* e para Campo Grande *News* utilizou-se o *Scrapy*. Como resultado, foram obtidas 55476 notícias do portal Campo Grande *News* e 5114 do portal *BBC News*. Foi possível obter a data de publicação, título, resumo e o corpo da notícia. Porém, por restrição da quantidade de memória disponível para os testes, foi utilizado apenas o resumo das notícias.

Após, foi realizado o processo de rotulação do *dataset*. Como se tratavam de muitos registros e o tempo disponível era exíguo, optou-se por uma estratégia automatizada, seguindo os passos:

Passo 1) Do conjunto de notícias obtidas foi necessária filtragem das notícias que realmente tinham relação com o domínio de interesse deste trabalho, violência contra a mulher. Para tanto, foi utilizada uma estratégia semelhante à nuvem de *tags*.

Nuvem de *tags* de notícias associadas a mulher:

*tags* mulher = [mulher, esposa, ex-esposa, ex esposa, tia, vó, avó, sogra, madrasta, sobrinha, noiva, filha, menina, enteada, ex-mulher, mãe, ex mulher, ex-noiva, ex noiva, namorada, ex-namorada, ex namorada, menina, adolescente, jovem, idosa, funcionaria,... ]

Foram mantidas apenas as notícias associadas à nuvem de *tags*, as outras foram descartadas. Com isso, restaram 3368 notícias oriundas do Campo Grande *News* e 319 notícias vindas do *BBC News*.

Passo 2) O próximo passo foi classificar as notícias que tratavam de casos de violência, assim, mais uma vez o processo foi automatizado por meio de uma segunda *tag* de palavras mais frequentes em tais situações. E as 3687

---

<sup>5</sup> <https://www.campograndenews.com.br/>

notícias foram avaliadas.

Nuvem de *tags* de notícias associadas à violência contra a mulher:

*tags violencia mulher = [violentada, estuprada, agredida, assassinada, baleada, esfaqueada, espancada, abusada, maltratada, ameaçada, estrangulada, socorrida, ajudada, medida protetiva, cárcere, maria da penha, vítima, fim do relacionamento, desrespeito, ciume, separação, amante, traição, consensual, ...]*

Como resultado, obtivemos: Campo Grande *News*: 2607 notícias neutras e 761 negativas e BBC *News*: 283 notícias neutras e 36 negativas. Por notícias negativas este trabalho entende todas as notícias associadas à nuvem de *tag tags mulher* e também à nuvem de *tag tags violencia mulher*.

As notícias e seus respectivos rótulos foram gravados em arquivo *csv* para compor a base de treinamento, validação e teste.

## 5. Configurações do modelo

Após estudos sobre processamento de textos, optou-se pelo modelo BERT (*Bidirectional Encoder Representations from Transformers*) (Devlin et al., 2019) uma vez que ele é citado como estado da arte em diversos contextos. Considerando, ainda, que o trabalho se vale de um conjunto de notícias no idioma português, foi utilizado o modelo Bertimbau, por já ter sido pré-treinado no idioma desejado (Souza et al., 2020).

Para realização dos testes, utilizou-se o repositório de modelos de Hugging-face<sup>6</sup>, cujo pré-treinamento foi feito utilizando BrWaC (*Brazilian Web as Corpus*) (Wagner et al., 2018). Dada a limitação de recursos computacionais para os testes, o trabalho optou pelo uso da versão *base*, com 12 *layers* e 110M parâmetros. Na saída do BERT foi conectado um classificador empregando *softmax*. Para a função de *loss* foi utilizada a *Cross Entropy*.

---

<sup>6</sup> <https://huggingface.co/neuralmind/bert-base-portuguese-cased>



## 6. Etapa de treinamento

Por se tratar de um conjunto de dados pequeno, seria importante o uso de alguma estratégia de *Data Augmentation*. Contudo, tal estratégia não foi empregada por ser necessário tempo para entendimento de diferentes bibliotecas disponíveis para este propósito, como *imgaug*<sup>7</sup>, *nlpaug*<sup>8</sup>, dentre outras.

Algumas estratégias de *fine tuning* foram utilizadas no modelo, envolvendo aplicação de gradiente descendente na atualização dos pesos de todas as camadas, e não apenas na camada de classificação conectada à saída do BERT. Também foi configurado o uso do *scheduler* para flexibilizar a taxa de aprendizagem e, assim, tentar impedir que o processo de otimização fique preso em mínimos locais.

Partindo destas configurações básicas, foram executados alguns treinamentos, sendo o modelo utilizado apresentado na Figura 1 e os dados obtidos mostrados na Tabela 1.

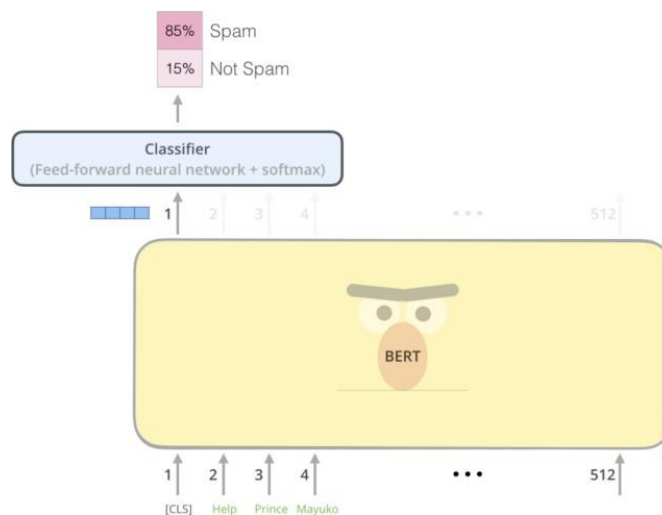


Figura 1 - Representação do modelo utilizado (OpenAI, 2021)

Tabela 1 – Dados do Treinamento

Tamanho <i>Dataset</i>	Total Neutro	Total Negativo	% <i>Train</i>	% Val	% <i>Test</i>	<i>Loss train</i>	<i>Loss val</i>	<i>Loss test</i>
1594	797	797	70%	20%	10%	0.6755	0,6935	0,7022

7 <https://github.com/aleju/imgaug>

8 <https://github.com/makcedward/nlpaug>

3687	2890	797	70%	20%	10%	0,8684	1,1143	0,8396
3687	2890	797	70%	20%	10%	0,6736	0,5490	0,6519
3687	2890	797	70%	20%	10%	0,6490	0,5150	0,6639

No primeiro teste, o *Dataset* fornecido como entrada possuía a mesma quantidade de notícias classificadas como Neutro e classificadas como Negativo (797 ocorrências de cada classe). Nos demais testes, essa equiparação não foi realizada, sendo fornecidas 2890 ocorrências da classe Neutro e 797 ocorrências da classe Negativo. O treinamento foi realizado usando ambiente de execução com uma GPU, disponibilizado no *Google Colab*.

## 7. Resultados

Todos os experimentos utilizaram 40 épocas com 200 passos cada. O processo de validação ao longo do treinamento compreendeu 50 passos por época. O objetivo era medir o impacto que determinadas configurações teriam no resultado do modelo treinado. Na Figura 2 é possível ver o valor da função *loss* do conjunto de treinamento e do conjunto de validação ao longo das épocas.

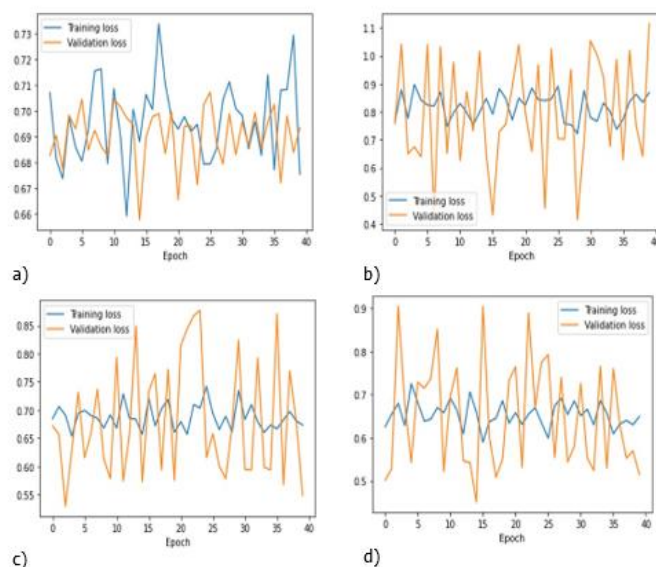


Figura 2 - Resultados de treinamentos realizados

A Figura 2 mostra resultados obtidos com treinamentos a partir de modelo pré-treinado com um *corpus* em língua portuguesa. Como as imagens

mostram claramente, o modelo não está convergindo em nenhum dos experimentos. O valor da função *loss* oscilou muito, o que leva à baixa acurácia nos testes. O modelo BERT é considerado bastante robusto e uma vez que se usa transferência de aprendizagem para inicialização de pesos, as respostas estão em torno de 0.8, conforme documentação de resultados do BERTimbau<sup>9</sup>.

## CONCLUSÃO

O presente trabalho teve como objetivo estudar estratégias para classificação de notícias através da Análise de Sentimento, com o objetivo de quantificar os casos envolvendo violência contra a mulher divulgados em jornais na web.

Para tanto, foi utilizado o modelo BERT, na sua versão pré-treinada no idioma português, chamado de BERTimbau. Com isso, a transferência de aprendizagem foi empregada, restando para trabalhos futuros avançar em possibilidades de aumento de dados.

Nos quatro treinamentos realizados, os resultados obtidos ficaram muito abaixo do relatado em *benchmarks* já publicados. Considerando que foram utilizados recursos de configuração convencionais e bem documentados<sup>10</sup>, é natural considerar como importante analisar a qualidade do *dataset* utilizado, tanto em qualidade das notícias quanto em quantidade. Assim, é inevitável que em trabalhos futuros o mesmo procedimento seja repetido com um *dataset* maior.

Importante considerar a manipulação de outros parâmetros no treinamento, além dos aqui descritos, com o objetivo de melhorar os resultados. Não deixando de destacar que o desbalanceamento do número de representantes de cada classe é um problema que pode criar viés no modelo, também é necessário investigar o método *Class Balanced Loss(CB)* e sua implementação no modelo BERT, o qual se propõe a contornar este problema.

---

9 <https://github.com/neuralmind-ai/portuguese-bert/>

10 <https://huggingface.co/transformers/v3.3.1/training.html>

Por fim, outras arquiteturas podem também ser utilizadas, para medir os resultados obtidos com o processo de classificação final.

Apesar dos resultados obtidos estarem aquém dos desejados, pensa-se que este processo, após um refinamento maior, irá auxiliar muito na identificação dos casos de violência contra a mulher.

## REFERÊNCIAS

AGÊNCIA Senado. DataSenado divulga pesquisa de violência contra a mulher nos estados e no DF. *Agência Senado*, 2024. Disponível em: <https://www12.senado.leg.br/noticias/materias/2024/02/28/datasenado-divulga-pesquisa-de-violencia-contra-a-mulher-nos-estados-e-no-df>. Acesso em: 08 de ago. de 2024.

CAMPOS, Ana Cristina. A cada 24 horas, ao menos oito mulheres são vítimas de violência. *Agência Senado*, 2024. Disponível em: <<https://agenciabrasil.ebc.com.br/direitos-humanos/noticia/2024-03/cada-24-horas-ao-menos-oito-mulheres-s%C3%A3o-vitimas-de-violencia>>. Acesso em: 08 de ago. de 2024.

DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton *et al.* Bert: Pre-training of deep bidirectional transformers for language understanding. *Cornell University*, 2019. Disponível em: <<https://arxiv.org/abs/1810.04805>>. Acesso em: 12 de mai. De 2024.

FACELI, Katti; LORENA, Ana Carolina; GAMA, João *et al.* *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. Rio de Janeiro: Livros Técnicos e Científicos, 2011.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *DeepLearning*. Cambridge: The MIT Press, 2016.

LIU, Bing. *Sentiment analysis and opinion mining* (Synthesis lectures on human language Technologies). Berlim: Springer, 2012.

OPENAI GPT-2: Compreendendo a geração de linguagem por meio da visualização. *ICHI.PRO*, 2021. Disponível em: <<https://ichi.pro/pt/openai-gpt-2-compreendendo-a-geracao-de-linguagem-por-meio-da-visualizacao-143292834755320>>. Acesso em: 05 de mai. 2024.

SILVA, Leandro Pereira da. *Leannet*: Uma arquitetura que utiliza o contexto da

cena para melhorar o reconhecimento de objetos. Dissertação (Mestrado em Ciência da Computação) – Escola Politécnica, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2018. Disponível em: <[https://tede2.pucrs.br/tede2/bitstream/tede/8168/2/DIS\\_LEANDRO\\_PEREIRA\\_DA\\_SILVA\\_COMPLETO.pdf](https://tede2.pucrs.br/tede2/bitstream/tede/8168/2/DIS_LEANDRO_PEREIRA_DA_SILVA_COMPLETO.pdf)>. Acesso em 08 de mai. 2024.

SOUZA, Fábio; NOGUEIRA, Rodrigo; LOTUFO, Roberto de Alencar. BERTimbau: pretrained BERT models for Brazilian Portuguese. *ResearchGate*, 2020. Disponível em:<[https://www.researchgate.net/publication/345395208\\_BERTimbau\\_Pretrained\\_BERT\\_Models\\_for\\_Brazilian\\_Portuguese](https://www.researchgate.net/publication/345395208_BERTimbau_Pretrained_BERT_Models_for_Brazilian_Portuguese)>. Acesso em 02 de mai. 2024.

VIOLÊNCIA Contra a Mulher. *nãosecale*, 2024. Disponível em: <<https://www.naosecale.ms.gov.br/violencia-contra-a-mulher/>>. Acesso em: 12 de jun. de 2024.

WAGNER, Jorge; WILKENS, Rodrigo Souza; IDIART, Marco; VILLAVICENCIO, Aline. The brWAC Corpus: A new open resource for brazilian portuguese. *ResearchGate*, 2018. Disponível em:<[https://www.researchgate.net/publication/326303825\\_The\\_brWaC\\_Corpus\\_A\\_New\\_Open\\_Resource\\_for\\_Brazilian\\_Portuguese](https://www.researchgate.net/publication/326303825_The_brWaC_Corpus_A_New_Open_Resource_for_Brazilian_Portuguese)>. Acesso em 02 de jun. 2024.