

UNIVERSIDADE FEDERAL DE ITAJUBÁ - UNIFEI
TRABALHO FINAL DE GRADUAÇÃO EM
CIÊNCIA DA COMPUTAÇÃO

**Comparação de modelos de Deep Learning na
detecção de comentários contendo discurso de ódio
na Internet**

Rafael Greca Vieira

1 de agosto de 2021

Itajubá

UNIVERSIDADE FEDERAL DE ITAJUBÁ - UNIFEI
TRABALHO FINAL DE GRADUAÇÃO EM
CIÊNCIA DA COMPUTAÇÃO

Rafael Greca Vieira

**Comparação de modelos de Deep Learning na
detecção de comentários contendo discurso de ódio
na Internet**

Monografia apresentada como trabalho final de graduação,
requisito parcial para obtenção do título de Bacharel em Ciência
da Computação, sob orientação da Profa. Dra. Isabela Neves
Drummond.

Orientador: Profa. Dra. Isabela Neves Drummond

1 de agosto de 2021

Itajubá

Agradecimentos

Os agradecimentos devem ocupar uma única página.

"Todos os modelos estão errados, mas alguns são úteis."
(George Box)

Resumo

O resumo deve ter no máximo 500 palavras e deve ocupar uma única página.

Palavras-chaves: palavra1; palavra2; palavra3; palavra5; palavra6.

Abstract

The abstract must have at most 500 words fit in a single page.

Key-words: word1; word2; word3; word4; word5.

Lista de ilustrações

Figura 1 – Perceptron.	5
Figura 2 – Rede Neural Artificial Multicamada	6
Figura 3 – Rede Neural Convolucional. Fonte: Adaptado de Phung, Rhee, et al. (2019)	6
Figura 4 – Primeiro passo na aplicação do filtro. Fonte: Kim (2017)	7
Figura 5 – Segundo passo na aplicação do filtro. Fonte: Kim (2017)	7
Figura 6 – Fluxograma das etapas do trabalho.	10
Figura 7 – Exemplos de mensagens da base de dados.	10
Figura 8 – Distribuição dos dados.	11
Figura 9 – Nova distribuição dos dados.	12

Lista de tabelas

Tabela 1 – Operações da EDA. Fonte: Adaptado de Wei e Zou (2019)	12
--	----

Lista de abreviaturas e siglas

EDA	Easy Data Augmentation
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
RNR	Rede Neural Recorrente
LSTM	Long-Short Term Memory
GRU	Gated Recurrent Unit
PLN	Processamento de Linguagem Natural
VP	Verdadeiro Positivo
FP	Falso Positivo
FN	Falso Negativo

Sumário

1	INTRODUÇÃO	1
2	FUNDAMENTAÇÃO TEÓRICA	3
2.1	Análise de Sentimentos	3
2.2	Aprendizado de Máquina	4
2.3	Aprendizado Profundo	4
2.3.1	Redes Neurais Artificiais	5
2.3.2	Redes Neurais Convolucionais	6
2.3.3	Redes Neurais Recorrentes	8
2.3.4	Long Short Term Memory	8
3	TRABALHOS RELACIONADOS	9
4	METODOLOGIA	10
4.1	Base de Dados	10
4.2	Tratamento dos Dados	11
4.3	Balanceamento dos Dados	11
4.4	Embedding	12
4.5	Modelos de Classificação	12
4.6	Ambiente de Execução	12
4.7	Avaliação dos Modelos	13
5	EXPERIMENTOS E RESULTADOS	14
5.1	Rede Neural Convolucional	14
5.2	Rede Neural Recorrente	14
5.3	Long Short Term Memory	14
5.4	Considerações Finais	14
6	CONCLUSÃO	15
	Referências	16

1 Introdução

Com a grande evolução da internet nos anos 90, não demorou muito para que as primeiras redes sociais começassem a surgir. Porém, sua popularização se deu apenas depois dos anos 2000 com a criação de algumas plataformas como: MySpace, Twitter e o Facebook. Nelas era possível se conectar com pessoas diferentes, postar fotos, vídeos, publicações e até conversar, em tempo real, com outra pessoa através do computador.

Essa popularização teve como consequência um aumento exponencial do volume de dados, de vários tipos, disponíveis e gerados todos os dias, tornando indispensável uma evolução computacional para que fosse possível lidar com a grande quantidade de dados. Este trabalho final de graduação tem como foco dados em formato de texto. Eles são trabalhados na subárea de inteligência artificial chamada de Processamento de Linguagem Natural e que tem como objetivo fazer com que o computador consiga compreender as palavras utilizadas nos mais diversos idiomas Chopra, Prashar, e Sain (2013).

Um dos principais campos de estudo, e que será abordado neste trabalho, na área de PLN é a análise de sentimentos, também chamada de mineração de opinião. Descrevendo de uma maneira bem sucinta, essa área de estudo tem como principal finalidade analisar a opinião humana, extraída de comentários publicados nas redes sociais, fóruns de discussão e sites de venda, seus sentimentos e emoções sobre um determinado assunto ou produto Torfi, Shirvani, Keneshloo, Tavaf, e Fox (2021).

O objetivo desse trabalho final de graduação é realizar uma comparação a respeito da utilização de técnicas de tratamento e balanceamento de dados entre algoritmos de aprendizado profundo voltado para a área de análise de sentimentos, mais especificamente na detecção de comentários na internet que podem ou não conter discurso de ódio.

O trabalho se concentra na comparação dos modelos de Rede Neural Convolucional, Rede Neural Recorrente e *Long-Short Term Memory*, em relação a acurácia dos resultados obtidos e ao tempo total de execução.

Essa monografia possui a seguinte estrutura:

- Capítulo 1: Introdução a respeito do que será abordado no trabalho;
- Capítulo 2: Serão apresentados os conceitos teóricos a respeito do campo de estudo escolhido, análise de sentimentos, as área de aprendizado de máquina e aprendizado profundo e, por fim, dos algoritmos escolhidos;
- Capítulo 3: Serão apresentados outros trabalhos relacionados com o tema escolhido, a base de dados que está sendo trabalhada ou com os modelos de aprendizado profundo

escolhidos;

- Capítulo 4: Será apresentada a metodologia escolhida para esse trabalho, detalhando qual a base de dados utilizada, qual o tratamento realizado nos dados, como é feito o balanceamento dos dados, qual o embedding escolhido, informações do ambiente onde será executado os testes e como os modelos desenvolvidos serão avaliados;
- Capítulo 5: Serão apresentados os experimentos e resultados obtidos para cada um dos algoritmos de aprendizado profundo escolhidos;
- Capítulo 6: Apresentação da conclusão do trabalho e possíveis trabalhos futuros.

2 Fundamentação Teórica

Este capítulo tem como finalidade descrever os conceitos mais importantes e fundamentais sobre os tópicos que serão abordados no trabalho. Nele será apresentado a respeito de análise de sentimentos, depois sobre aprendizado de máquina e aprendizado profundo. Dentro da seção sobre aprendizado profundo será discutido sobre Rede Neural Artificial, Rede Neural Convolucional, Rede Neural Recorrente e *Long Short Term Memory*.

2.1 Análise de Sentimentos

Esse campo de estudo é um dos mais ativos da área de processamento de linguagem natural e visa analisar os sentimentos, opiniões e emoções das pessoas através de um texto escrito Liu (2012).

A análise de sentimentos, ou também chamada de mineração de opinião, vem ganhando ainda mais destaque nos últimos anos. Isso se deve, principalmente, ao grande crescimento das redes sociais e blogs pessoais. Nelas, como Facebook e Twitter, os usuários são livres para falarem sobre o que quiserem, seja a respeito de uma opinião política, avaliação de um produto ou empresa, opinião sobre um determinado assunto, dentre outras inúmeras possibilidades Kouloumpis, Wilson, e Moore (2011).

Além das redes sociais e blogs, os fóruns de discussão, onde milhares de pessoas reúnem com o mesmo propósito de discutir e opinar sobre um determinado assunto, e sites de venda, onde os clientes podem avaliar os produtos e a qualidade do serviço de uma determinada empresa, também proporcionam uma grande visibilidade nesse campo de estudo.

A análise de sentimento pode ser tratada como um problema de classificação, pois almeja detectar qual a polaridade e o sentimento contido em um determinado texto. Com o avanço dos estudos nessa área, essa classificação não está mais limitada em ser apenas um sentimento positivo ou negativo, mas também é possível detectar as emoções em diferentes tópicos Ramadhani e Goo (2017).

Por ser uma área muito abrangente, ela possui diversas aplicações e em diversas áreas. Algumas delas são: análise sobre investimentos no setor financeiro Feldman (2013), reputação de uma marca nas redes sociais Feldman (2013), avaliação dos clientes sobre um determinado produto ou serviço Feldman (2013), análise de publicações de perfis pessoais e coleta de feedbacks em chatbots.

2.2 Aprendizado de Máquina

Diversos projetos na área de inteligência artificial utilizam regras de inferência lógica para tomar suas decisões. Esses sistemas são chamados de base de conhecimento. A grande dependência desse tipo de sistema ocasionou na criação de um sistema que possui capacidade de adquirir seu próprio conhecimento por meio da extração de padrões em dados bruto. Essa capacidade é chamada de aprendizado de máquina Goodfellow, Bengio, e Courville (2016).

A etapa de aprendizagem é realizado durante o processo de treinamento no modelo. Onde, além de produzir os resultados desejados para os dados utilizados no treinamento, ele aprende a generalizar visando produzir os resultados esperados para dados nunca antes vistos El Naqa e Murphy (2015).

O modelo de aprendizado de máquina pode ser categorizado levando em consideração o seu processo de aprendizagem. Essas categorias são: aprendizagem supervisionada, não-supervisionada, semi-supervisionada e por reforço.

Os modelos supervisionados utilizam dados de treinamento, junto com as suas respectivas classes, no processo de aprendizagem. Os não-supervisionados aprendem através do reconhecimento de padrões ocultos nos dados. Os semi-supervisionados possuem os dois tipos de dados El Naqa e Murphy (2015). Os modelos de aprendizado por reforço não recebem ordens do que fazer e, através de reforço, aprende o que fazer visando maximizar uma determinada recompensa Sutton e Barto (2018).

Além da rotulação dos modelos em relação a sua aprendizagem, eles também podem ser classificados de acordo com o tipo de tarefa que desejam solucionar. Ela pode ser de classificação ou regressão.

Para a tarefa de classificação, o resultado do modelo será um valor discreto e que representa uma determinada classe dentro de um determinado conjunto de possibilidades. Já para a de regressão, o resultado do modelo será um valor contínuo e de infinitas possibilidades Rätsch (2004).

Existem diversos algoritmos diferentes de aprendizado de máquina que são aplicados nos mais diversos tipos de dados, incluindo aqueles no formato de texto. Kajla, Hooda, Saini, et al. (2020) realizam uma comparação utilizando seis desses algoritmos na classificação de comentários tóxicos. Eles são: Regressão Linear, Naive Bayes, Árvore de Decisão, Florestas Randômicas, *K-Nearest Neighbors* e *Support Vector Machine*. O que teve o melhor resultado, em termos de desempenho, foi o algoritmo de Regressão Linear.

2.3 Aprendizado Profundo

Explicar aprendizado profundo.

2.3.1 Redes Neurais Artificiais

A Rede Neural Artificial (RNA) é inspirada nos modelos biológicos humanos. A rede neural é composta pelos seguintes elementos: neurônio artificial, que representam os neurônios biológicos; os dados de entrada, que representam os sinais que são recebidos pelos dendritos; pesos, que representam as forças dos dados de entrada; a função de ativação, que representa uma função que irá ou não ativar o neurônio (característica não linear); a saída, representada pela impulso do neurônio Abraham (2005); Gershenson (2003).

A saída é computada através da soma ponderada das entradas com seus respectivos pesos. Esse valor será passado para a função de ativação. Por fim, o resultado dessa função será utilizado em uma outra função de ativação que varia dependendo do objetivo do modelo. Quando não é utilizada a segunda função de ativação (o valor final é igual ao valor resultante da primeira função de ativação), o neurônio pode ser chamado de linear Gershenson (2003).

Um dos primeiros modelos, e o mais conhecido, de uma rede neural desenvolvido foi o *Perceptron*, composto por uma camada que possui apenas um neurônio e que pode ser visualizado na figura 1.

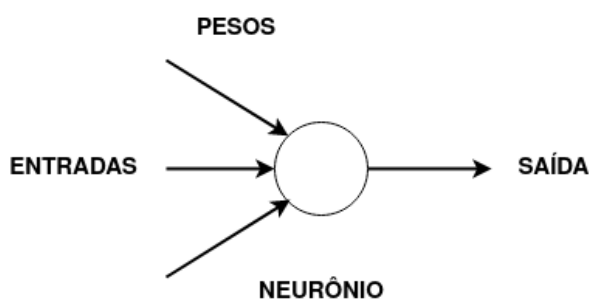


Figura 1 – Perceptron.

Uma evolução do modelo anterior, e consequentemente mais robusta, é a RNA multicamadas. Ela é composta por três camadas: entrada, escondida e saída. A camada de entrada é por onde os dados entram na rede neural, portanto é a primeira da rede. A camada escondida vem logo em seguida. Por fim, a última camada é a de saída e irá gerar o resultado obtido para os determinados dados Abraham (2005), como está representado na figura 2. Quando a rede possui mais de uma camada escondida ela é chamada de rede neural profunda.

A rede neural artificial é o exemplo mais conhecido de uma rede neural *feed-forward*. Em uma rede *feed-forward* as informações percorrem a rede neural em somente uma direção. As setas representadas na figura 2 correspondem ao fluxo das informações na rede. Elas entram pela camada de entrada, depois seguem para a(s) camada(s) escondida(s) até chegar na de saída.

O processo de aprendizado de uma rede neural se dá pela atualização dos pesos durante a fase de treinamento por meio de um processo supervisionado. O erro é calculado pela diferença da saída esperada com a saída obtida da rede. O algoritmo de *backpropagation* irá atualizar os

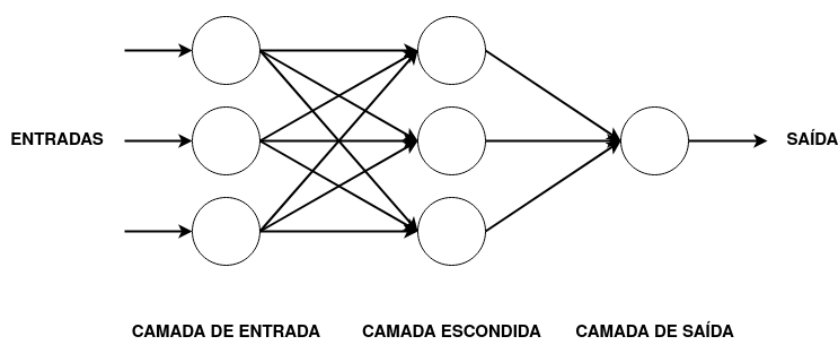


Figura 2 – Rede Neural Artificial Multicamada

pesos da RNA, utilizando os erros, percorrendo o caminho oposto, ou seja, da camada saída até a de entrada Werbos (1990).

2.3.2 Redes Neurais Convolucionais

A Rede Neural Convolucional (RNC), ou Rede Convolucional, possui a mesma ideia e funcionamento que a RNA. A grande diferença entre as duas é que a RNC é composta por camadas convolucionais ao invés da tradicional camada escondida da RNA Goodfellow et al. (2016). Um exemplo de arquitetura dessa rede pode ser vista na figura 3.

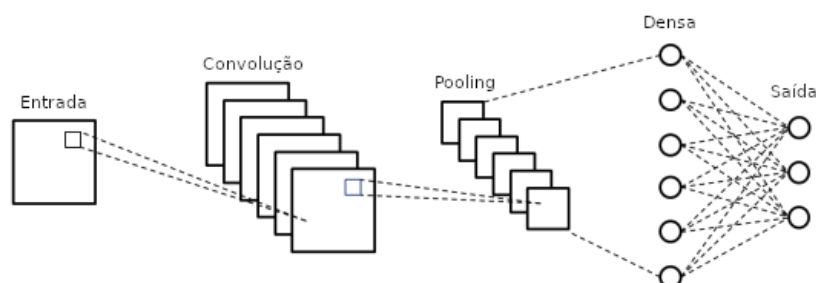


Figura 3 – Rede Neural Convolucional. Fonte: Adaptado de Phung et al. (2019)

Diferente de uma RNA comum, a RNC recebe uma entrada que está no formato de uma matriz numérica, contendo pelo menos uma dimensão, como uma imagem, texto, *time-series* ou áudio Goodfellow et al. (2016). A imagem é representada pelos valores dos seus pixels, o texto pelos valores numéricos das suas palavras através de um *embedding*, *time-series* com suas informações em relação a um determinado período de tempo sequencial e áudio com o timbre e frequência da voz, por exemplo.

Além das camadas de entrada e saída, a arquitetura da RNC é composta de outras três camadas: convolucional, *pooling* e densa.

A camada convolucional é uma das principais camadas responsáveis pela extração de características dos dados. Ela utiliza filtros, também chamados de *kernels*, de pequena dimensão e que são aprendidos ao longo do treinamento da rede neural O'Shea e Nash (2015). Os filtros

serão deslocados pelos dados, de acordo com um determinado valor de *stride*, a operação de convolução será realizada entre essas duas matrizes e o resultado será utilizado para criar um novo dado. A saída dessa camada será um mapa de informações do mesmo tamanho que a quantidade de filtros utilizados Kim (2017).

1	1	1	3
4	6	4	8
30	0	1	5
0	2	2	4

 $\odot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$

7			

Figura 4 – Primeiro passo na aplicação do filtro. Fonte: Kim (2017)

Na figura 4 a entrada é representada pela primeira matriz, a operação de convolução pelo símbolo do asterisco, o filtro pela segunda matriz e, por fim, o novo dado gerado pela terceira matriz. Será selecionada uma pequena área da matriz, cuja a dimensão é a mesma que a do filtro, será feita a operação de convolução entre eles e o resultado será armazenado em uma nova matriz. Na figura 5 foi selecionada outra área deslocando um espaço para o lado, *stride* de valor igual a 1 (um), e depois realizando as mesmas etapas mencionadas anteriormente. Isso será feito até que toda a área da entrada tenha sido utilizada, resultando em um novo mapa de informações. Caso o número de filtros utilizados seja 6 (seis), a saída da camada de convolução seria a mesma que está representada na figura 3.

1	1	1	3
4	6	4	8
30	0	1	5
0	2	2	4

 $\odot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$

7	5		

Figura 5 – Segundo passo na aplicação do filtro. Fonte: Kim (2017)

A camada *pooling* é a responsável por reduzir a dimensão da saída da camada anterior para as próximas. Isso garante que o modelo fique menos complexo, pois terá menos parâmetros para serem treinados, sem perder as informações mais relevantes O'Shea e Nash (2015). Porém, por mais que o conteúdo dessas informações sejam preservados, a sua informação espacial, como a posição original no dado, é perdida Albawi, Mohammed, e Al-Zawi (2017).

Os neurônios da camada densa, também chamada de totalmente conectada, possui uma estrutura idêntica a maneira como os neurônios da RNA estão conectados. Todos eles são conectados com os neurônios da camada anterior e posterior Albawi et al. (2017).

2.3.3 Redes Neurais Recorrentes

Este capítulo tem como finalidade...

2.3.4 Long Short Term Memory

Este capítulo tem como finalidade...

3 Trabalhos Relacionados

Ibrahim, Torki, e El-Makky (2018) utilizaram as técnicas de pré processamento e balanceamento de dados na detecção de comentários tóxicos em um conjunto de dados disponibilizados no Kaggle. Foi feita dois tipos de classificação: binária (possui toxicidade ou não) e multiplas categorias. Para cada uma das classificações foram utilizadas três modelos de aprendizado profundo: RNC, LSTM bidirecional e GRU bidirecional. O modelo que apresenta o melhor resultado em ambas classificações foi uma combinação dos três modelos citados anteriormente.

4 Metodologia

Este capítulo tem como objetivo descrever todas as etapas que serão executadas neste trabalho, desde a coleta da base de dados utilizada até a avaliação dos modelos escolhidos, conforme mostra o fluxograma da figura 6.

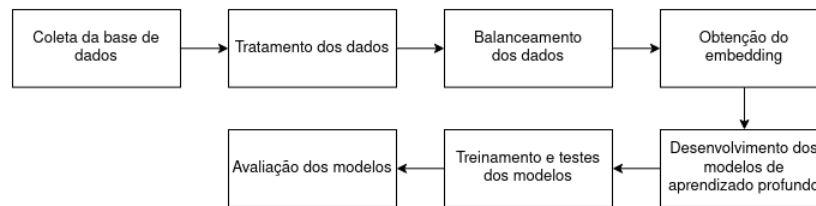


Figura 6 – Fluxograma das etapas do trabalho.

4.1 Base de Dados

A base de dados utilizada no trabalho se chama *Twitter Sentiment Analysis* e foi originalmente disponibilizada em uma competição no site Analytics Vidhya, que teve seu início no começo de 2018. Mais tarde, esse conjunto de dados também foi disponibilizada no site Kaggle.

A base é composta por publicações em inglês feitas por usuários do Twitter. Ela é dividida em dois arquivos: treinamento, que possui aproximadamente 32 mil *tweets* categorizados como contendo ou não discurso de ódio; teste, possui aproximadamente 17 mil *tweets*. A figura 7 mostra alguns dos *tweets* que estão contidos na base.

id	label	tweet
1	0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run
2	0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disappointed #getthankd
3	0	bihday your majesty
4	0	#model i love u take with u all the time in urõ±!!! õõõõõ!õ!õ!
5	0	factsguide: society now #motivation
6	0	[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo
7	0	@user camping tomorrow @user @user @user @user @user @user @user dannyâ!
8	0	the next school year is the year for exams.õ can't think about that õ #school #exams #hate #imagine #actorslife #revolutionschool #girl
9	0	we won!!! love the land!!! #allin #cavs #champions #cleveland #clevelandcavaliers â!
10	0	@user @user welcome here ! i'm it's so #gr8 !
11	0	â #ireland consumer price index (mom) climbed from previous 0.2% to 0.5% in may #blog #silver #gold #forex
12	0	we are so selfish. #orlando #standwithorlando #pulseshooting #orlandoshooting #biggerproblems #selfish #heabreaking #values #love #
13	0	i get to see my daddy today!! #80days #gettingfed
14	1	@user #cnn calls #michigan middle school 'build the wall' chant " #tcot
15	1	no comment! in #australia #opkillingbay #seashepherd #helpcovedolphins #thecove #helpcovedolphins
16	0	ouch...junior is angryõ#got7 #junior #yugyoem #omg
17	0	i am thankful for having a paner. #thankful #positive

Figura 7 – Exemplos de mensagens da base de dados.

Cada *tweet* terá um id, que é um número único de identificação, um booleano identificando se contém ou não discurso de ódio e, por fim, o seu conteúdo bruto. Na figura 8 é possível visualizar como os dados estão distribuídos dentro da base. Os *tweets* que não contém discurso

de ódio são a maioria, contendo quase 30 mil, enquanto os *tweets* que possuem discurso de ódio são apenas um pouco mais de 2 mil.

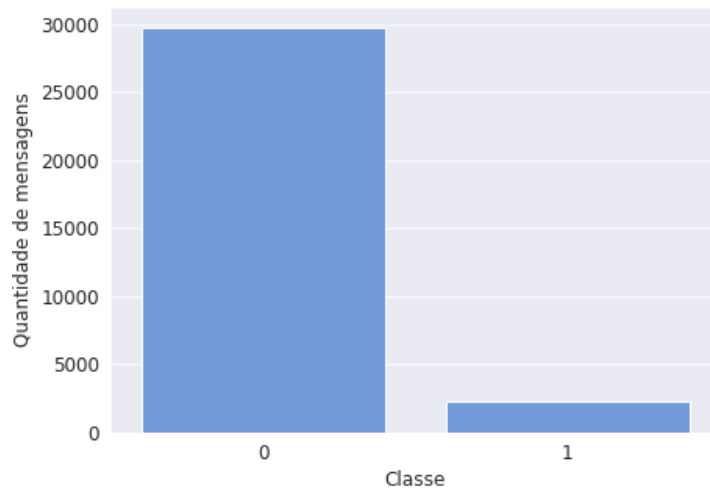


Figura 8 – Distribuição dos dados.

4.2 Tratamento dos Dados

Texto...

4.3 Balanceamento dos Dados

Na figura 6 é possível notar que a base de dados que está sendo utilizada é totalmente desbalanceada, já que a classe majoritária representa, aproximadamente, 93% da quantidade total de dados. A maioria dos modelos de aprendizado profundo atuais não levam em consideração o desbalanceamento dos dados. Essa distribuição não normal das classes pode afetar o aprendizado do modelo, pois a classificação será mais tendenciosa para a classe majoritária Wang et al. (2016).

Para a realização do balanceamento dos dados será utilizada a técnica EDA. Ela aprimora a performance em tarefas de classificação utilizando quatro operações: substituição por um sinônimo, inserção aleatória, troca aleatória e remoção aleatória Wei e Zou (2019). Essas operações estão representadas na tabela 1.

A técnica será utilizada para incrementar o volume da classe minoritária. Os valores escolhidos para os parâmetros foram os indicados pelos desenvolvedores do EDA. A figura 9 apresenta a nova distribuição dos dados após a realização do balanceamento. A classe majoritária agora representa cerca de 65% do total de dados e não mais 93%.

Tabela 1 – Operações da EDA. Fonte: Adaptado de Wei e Zou (2019)

Operação	Sentença
Nenhuma	Uma comédia humana triste e superior jogada fora nas estradas secundárias da vida.
Substituição Sinônimo	A Uma comédia humana lamentável e superior jogada fora nas estradas retrógradas da vida.
Inserção Aleatória	Uma comédia humana triste e superior jogada fora nas graça estradas secundárias da vida.
Troca Aleatória	Uma comédia humana triste e superior jogada fora nas secundárias estradas da vida.
Remoção Aleatória	Uma comédia humana triste e superior fora nas estradas da vida.

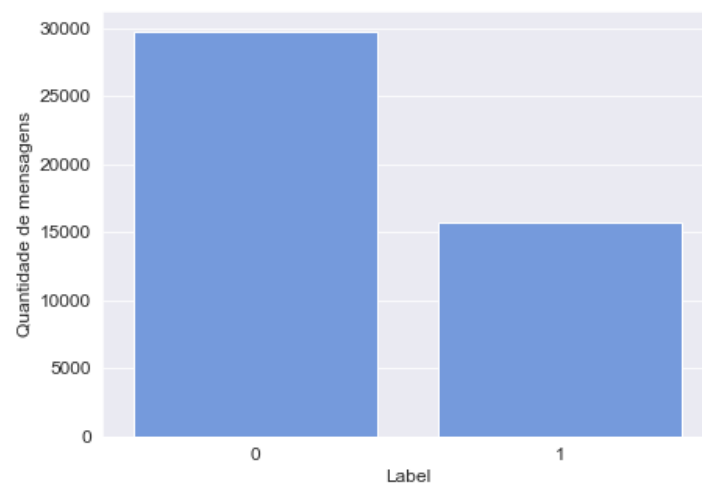


Figura 9 – Nova distribuição dos dados.

4.4 Embedding

Texto...

4.5 Modelos de Classificação

Texto...

4.6 Ambiente de Execução

Texto...

4.7 Avaliação dos Modelos

Os modelos serão avaliados utilizando a métrica adotada pela competição onde os dados se encontram: o *F1-Score*. Essa métrica é composta de outras duas métricas: *Precision* e *Recall*. O *Precision* é calculado utilizando Verdadeiro Positivo (VP) e Falso Positivo (FP), como mostra a equação 4.1. Já o *Recall* é calculado utilizando VP e Falso Negativo (FN), como mostra a equação 4.2.

$$Precision = \frac{VP}{VP + FP} \quad (4.1)$$

$$Recall = \frac{VP}{VP + FN} \quad (4.2)$$

A fórmula do F1-Score está representada na equação 4.3. Para a sua implementação será utilizada a biblioteca *scikit-learn* Pedregosa et al. (2011).

$$F1Score = \frac{2 * (Recall * Precision)}{Recall + Precision} \quad (4.3)$$

5 Experimentos e Resultados

5.1 Rede Neural Convolucional

Este capítulo tem como finalidade...

5.2 Rede Neural Recorrente

Este capítulo tem como finalidade...

5.3 Long Short Term Memory

Este capítulo tem como finalidade...

5.4 Considerações Finais

Texto..

6 Conclusão

Concluiu-se que...

Referências

- Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*.
- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (icet)* (pp. 1–6).
- Chopra, A., Prashar, A., & Sain, C. (2013). Natural language processing. *International journal of technology enhancements and emerging engineering research*, 1(4), 131–134.
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In *machine learning in radiation oncology* (pp. 3–11). Springer.
- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82–89.
- Gershenson, C. (2003). Artificial neural networks for beginners. *arXiv preprint cs/0308031*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (<http://www.deeplearningbook.org>)
- Ibrahim, M., Torki, M., & El-Makky, N. (2018). Imbalanced toxic comments classification using data augmentation and deep learning. In *2018 17th IEEE international conference on machine learning and applications (icmla)* (pp. 875–878).
- Kajla, H., Hooda, J., Saini, G., et al. (2020). Classification of online toxic comments using machine learning algorithms. In *2020 4th international conference on intelligent computing and control systems (iciccs)* (pp. 1119–1123).
- Kim, P. (2017). Convolutional neural network. In *Matlab deep learning* (pp. 121–147). Springer.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! In *Fifth international AAAI conference on weblogs and social media*.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1–167.
- O’Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12, 2825–2830.
- Phung, V. H., Rhee, E. J., et al. (2019). A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. *Applied Sciences*, 9(21), 4500.
- Ramadhani, A. M., & Goo, H. S. (2017). Twitter sentiment analysis using deep learning methods. In *2017 7th international annual engineering seminar (inaes)* (pp. 1–4).

- Rätsch, G. (2004). A brief introduction into machine learning. *Friedrich Miescher Laboratory of the Max Planck Society*.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., & Fox, E. A. (2021). *Natural language processing advancements by deep learning: A survey*.
- Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., & Kennedy, P. J. (2016). Training deep neural networks on imbalanced data sets. In *2016 international joint conference on neural networks (ijcnn)* (pp. 4368–4374).
- Wei, J., & Zou, K. (2019, November). EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 6383–6389). Hong Kong, China: Association for Computational Linguistics. Retrieved from <https://www.aclweb.org/anthology/D19-1670>
- Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550–1560.