

# Utilização de técnicas de aprendizado de máquina supervisionado para detecção de Fake News

Renan Rocha de Andrade

<sup>1</sup>Curso de Ciência da Computação– Universidade do Sul de Santa Catarina (UNISUL) – Tubarão – SC – Brasil

andrade.renan@unisul.br

**Abstract.** *Fake news is an issue of increasing concern in recent times due to its impact. Certain agencies of investigative journalism carry out a fight against fake news, but as they depend on human effort, this process is not so efficient. In this work it is made the use of natural language processing techniques and supervised machine learning, in order to generate a news classification model, to assist in the process of identifying fake news, in a way that it creates a good level of confidence, in an appropriate computational time. In the implementation testing, it was possible to achieve up to 91% accuracy.*

**Resumo.** *Notícias falsas são um problema de preocupação crescente nos últimos tempos, devido ao seu impacto causado. Certas agências de jornalismo investigativo realizam esse combate às notícias falsas, porém como dependem do esforço humano, esse processo não é tão eficiente. Neste trabalho é realizado o uso de técnicas de processamento de linguagem natural e aprendizado de máquina supervisionado, de forma a produzir um modelo de classificação de notícias, para auxiliar no processo de identificação de notícias falsas, de maneira que produza um bom nível de confiança, em tempo computacional adequado. Nos testes da implementação, foi possível alcançar até 91% de taxa de acerto.*

## 1. Introdução

Atualmente, com os avanços das tecnologias digitais, indivíduos tem acesso maior e muito mais simplificado a manchetes e notícias, desprendendo-se das mídias tradicionais como televisão, jornal e rádio, e aproximando-se de mídias alternativas disponíveis na internet, em blogs e redes sociais.

No entanto, nem toda informação que circula amplamente nessas mídias alternativas é realmente verdadeira.

Segundo The Guardian (2016), *fake news* referem-se ao tipo de notícias que apresentam informações falsas, que tem forte apelo para o emocional de seus leitores com o objetivo de aumentar taxas de acessos em sites de internet, influenciar opinião pública de maneira contrária ou favorável a determinado assunto.

Segundo Zanatta (2018), pesquisador da Universidade de São Paulo, em artigo do portal Outras Palavras, quem cria *fake news* promove mentira e manipula cidadãos em torno de interesses particulares e desonestos.

Em estudo divulgado pelo site G1 (2018), uma notícia falsa tem uma probabilidade de ser compartilhada de até 70% maior que uma notícia verdadeira.

O problema da fabricação de notícias falsas e sua divulgação vem ganhando cada vez mais destaque em escala mundial, principalmente com as eleições presidenciais dos Estados Unidos em 2016 e as eleições no Brasil em 2018.

Diante disso, surgiram certas agências com o intuito de identificar e combater a divulgação de notícias falsas. Portais de jornalismo investigativo tem o objetivo de analisar conteúdo divulgado na internet, principalmente em redes sociais, para desmentir notícias falsas. Porém, essas agências necessitam de verificação humana, o que dificulta muito o processo de combate a *fake news*.

Com o grande volume de dados produzidos, propõe-se desenvolver um modelo de aprendizado de máquina supervisionado, que pode ser treinado a partir de notícias que já se sabe que são verdadeiras ou falsas, para realizar a classificação das notícias, de modo a diminuir o impacto negativo desse tipo de notícia na sociedade.

A segunda seção desse artigo apresenta alguns trabalhos correlatos que tratavam sobre classificação de textos, alguns com técnicas similares às aplicadas nesse trabalho. Na seção 3 temos uma contextualização com os conceitos principais das técnicas realizadas nesse trabalho. Na seção 4 há o detalhamento dos materiais utilizados no trabalho, assim como a descrição detalhada da metodologia. A seção 5 contém os resultados alcançados no trabalho. Por fim a seção 6 contém uma conclusão do trabalho realizado.

## 2. Trabalhos Relacionados

Dado o problema de classificação de notícias em verdadeiras e falsas, nota-se que o mesmo se trata de um problema de classificação de textos, tratado na literatura em trabalhos publicados em diferentes domínios.

Diante disso, é possível descrever os trabalhos sobre duas classes referentes a sua representação: representações clássicas e representações atuais

### 2.1. Representações clássicas

É dito como representações clássicas, pois grande parte dos artigos utiliza essa abordagem baseada numa estrutura de matriz de termos e documentos (Rossi, 2015).

Monteiro (2018), no seu projeto de iniciação científica de detecção automática de notícias falsas em português, utilizou além de matriz termo-documento (chamado no artigo de bag of words), também utilizou POS tags (classes gramaticais), pausalidade, emotividade, e constatou que a matriz termo-documento foi a melhor representação, com os algoritmos de classificação testados: *naive bayes*, SVM<sup>1</sup>, árvore de decisão (J48), *random forest*. Chegou a resultados 88% a 89% de acurácia.

---

<sup>1</sup> Support Vector Machine (SVM) é um método de aprendizado supervisionado

Ahmed et al. (2017), no seu artigo de detecção de spams de opiniões e *fake news*, utiliza também essa representação juntamente com TF<sup>2</sup> e TF-IDF<sup>3</sup>, para definir uma relevância dos termos quanto sua frequência, e tokens que vão de 1-gram a até 4-grams. Além disso realizou uma comparação dos métodos de classificação SVM, kNN<sup>4</sup>, árvore de decisão, regressão logística.

Vale ressaltar nesse artigo é que na maioria dos casos o SVM teve melhores resultados, mas os atributos se diferenciaram nos *datasets*. Em alguns casos foi melhor utilizar 1-gram, em outros casos 2-grams, resultados variaram precisões de 70% até 92% dependendo muito de como estão representados os atributos. Foi possível perceber que acima de 2-grams, não há ganho de precisão de classificação.

## 2.2. Representações atuais

Trabalhos publicados em classificação de textos sobre representações alternativas à matriz termo-documento, utilizam representações mais atuais e seus algoritmos classificadores apresentaram resultados muitas vezes, superiores aos obtidos sobre as representações clássicas.

Rossi (2015), em sua tese de doutorado, propôs um modelo de classificação de textos baseado em redes/grafos. Para isso fez uma extensa avaliação sobre modelos de matriz termo-documento, para comparar com o modelo proposto baseado em grafos.

Essa representação se baseia em representar os termos e documentos por meio de um grafo. Em sua tese ele discorre sobre diversas maneiras de como poderia haver essa representação.

Na sua tese ele também propôs o algoritmo IMHN<sup>5</sup>, para realizar a classificação sobre esse modelo de representação, e para diversos *datasets* testados, o algoritmo desenvolvido na sua tese se saiu melhor.

Na comparação com outros algoritmos, vale constatar que o *naive bayes* teve a segunda melhor performance sobre os *datasets* testados por ele.

Lai et al. (2015), em seu artigo de redes neurais recorrentes convolucionais para classificação de texto, por meio de conjuntos pré-treinados de *word embeddings* (incorporação de palavras, em tradução livre), juntamente com redes neurais de *deep learning* para realizar essa tarefa de classificação.

Essa representação de *word embeddings* é muito poderosa para representar o contexto de palavras em textos, mostrar relações de similaridade e sentido e relações entre uma palavra e outra.

Nesse artigo, o autor utiliza um modelo de *word embeddings* pré-treinado, e a partir dele aplica funções no vetor de determinada palavra do documento para adquirir seu contexto. O contexto é representado por outro vetor.

---

<sup>2</sup> Term Frequency (TF) é uma técnica para determinar importância de um termo num texto

<sup>3</sup> Term Frequency Inverse Document frequency é uma técnica para determinar importância de um termo num texto

<sup>4</sup> K Nearest Neighbors (kNN) é um método de aprendizado supervisionado

<sup>5</sup> Inductive Model using Heterogeneous Networks é um método classificador de textos

A partir disso, essa estrutura é usada num modelo de *deep learning*, CNN (Convolutional Neural Network) ou RCNN (Recurrent Convolutional Neural Network), para obter a classificação, com precisões atingindo até 95%.

### 3. Contextualização

#### 3.1 Word Embedding

Segundo Karani (2018), *word embedding* é uma forma de representação de vocabulário. Trata-se de representar uma determinada palavra através de um vetor de números, sendo que essa representação leva em consideração os contextos das palavras.

Para obter essas representações são utilizadas redes neurais, que são treinadas em diversos textos, para no final produzir vetores “próximos” para palavras de um contexto comum e “distantes” para palavras em contextos diferentes.

#### 3.2 Redes Neurais

Segundo Vaz (2018), redes neurais artificiais são algoritmos inspirados pelo funcionamento do sistema nervoso central humano, especialmente às sinapses cerebrais. Tais algoritmos aprendem a realizar uma tarefa, seja de classificação ou regressão, através de exemplos.

Uma rede neural é construída a partir de uma série de camadas conectadas formadas por neurônios. Essas conexões podem transmitir “sinais” de um neurônio a outro, como as sinapses em um cérebro. Cada conexão possui uma matriz de pesos, estes que são ajustáveis durante a fase de aprendizado. Cada neurônio possui uma função de ativação, que leva um neurônio a propagar ou inibir o sinal (Vaz, 2018). A figura 1 representa esse funcionamento. Em *Input* temos a entrada de dados,  $W_1$ ,  $W_2$ ,  $W_o$ , possuem as matrizes de pesos,  $f_1$  corresponde às funções de ativação dos neurônios da *hidden layer 1*,  $f_2$  corresponde às funções de ativação dos neurônios da *hidden layer 2* e por fim *output layer* é a camada de neurônios de saída que fornecem o valor previsto a partir da entrada.

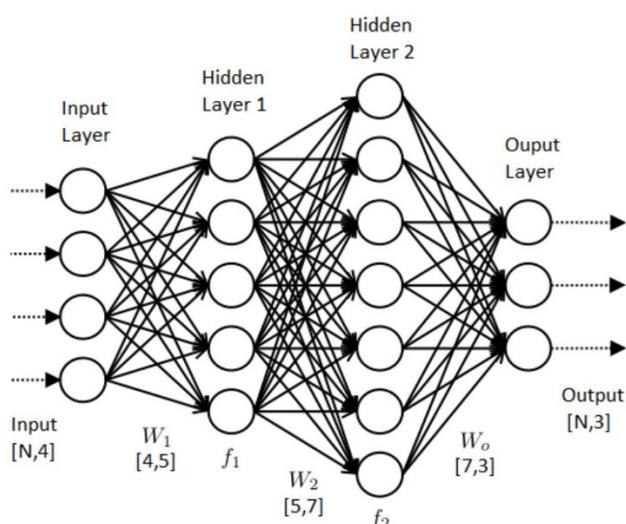


Figura 1: Exemplo de rede neural  
Fonte: DZone

Sua técnica é baseada em duas etapas: treinamento, na qual o modelo da rede neural é treinado a partir de um conjunto de dados de treino, durante um determinado

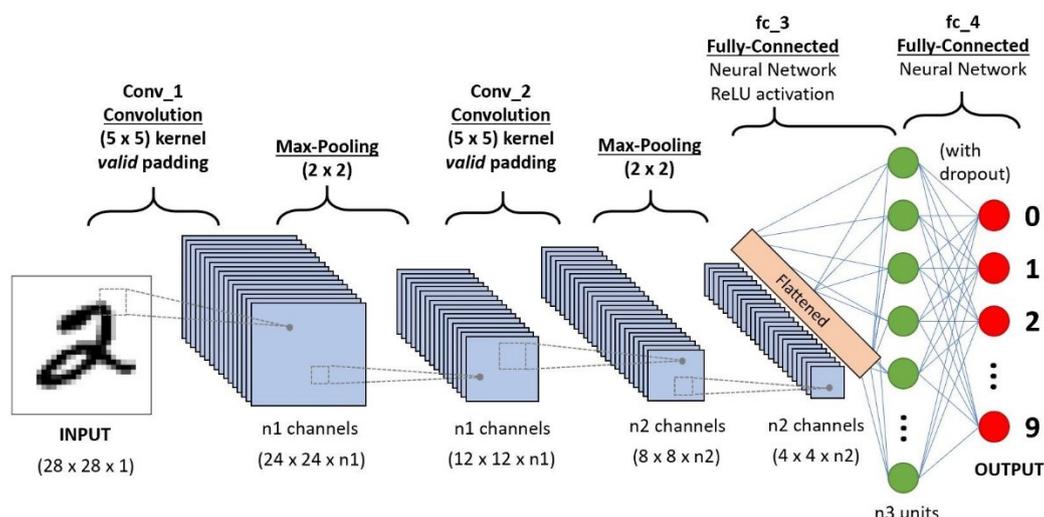
número de épocas. A partir desse treinamento, os valores da matriz de peso são ajustados para se obter uma saída mais próxima da desejada. Já a etapa de teste, não há ajuste de pesos, o modelo é executado num conjunto de dados de treinamento com o intuito de medir a performance do mesmo.

### 3.3 Deep Learning

*Deep learning* é um método de aprendizado de máquina, baseado em redes neurais. Esse método baseia-se no uso de diversas “camadas escondidas” (*hidden layers*) em uma rede neural, para criar o modelo. É por essa característica que se dá o nome de aprendizado profundo.

Segundo LeCun et al. (2015), técnicas convencionais de aprendizado de máquina possuíam limitações no processamento de dados naturais, necessitando de grande experiência sobre o domínio para realizar extração de atributos. *Deep learning* permite criar métodos de aprendizado de máquina com camadas capazes de aprender representações de dados e fazer transformações, com diferentes níveis de abstração.

Essa técnica está tendo crescente relevância dentro das aplicações de inteligência artificial e vem sendo usado em diversas áreas como, reconhecimento de voz, visão computacional, processamento de linguagem natural, análises toxicológicas, entre outros (LeCun et al., 2015). A figura 2 contém uma representação de uma rede *deep learning* para classificar dígitos manuscritos. *Conv\_1* e *Conv\_2* representam camadas convolucionais, que realizam filtros na matriz de entrada, as camadas de *max-pooling* obtém o valor máximo de uma janela deslizante através da saída das camadas convolucionais, *fc\_3* e *fc\_4* representam camadas com neurônios totalmente conectados com suas funções de ativação, e output que é a saída com os valores previstos.



**Figura 2: Exemplo de rede convolucional *deep learning***  
**Fonte: Towards Data Science**

De acordo LeCun et al. (2015) a perspectiva do *deep learning* no futuro é muito promissora, dado o aumento na quantidade de dados, poder computacional e pesquisas envolvendo essa metodologia.

### 3.4 Naïve Bayes

Segundo Becker (2019), Naïve Bayes é um algoritmo classificador probabilístico, baseado no teorema de Bayes de probabilidade condicional. É um algoritmo popular de aprendizado de máquina, muito usado na classificação de textos, como para diferenciar e-mails que são ou não *spam*.

O teorema de Bayes foi publicado pelo estatístico Thomas Bayes em 1763, que se trata da probabilidade condicional, este que é um dos conceitos mais importantes da teoria da probabilidade (Ross, 2010). Segundo Gut (2013), a probabilidade condicional é a medida de probabilidade ocorrer, dado que outro evento tenha ocorrido.

A equação do teorema de Bayes é dada por:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

em que  $P(A)$  corresponde a probabilidade do evento  $A$ ,  $P(B)$  corresponde a probabilidade do evento  $B$ ,  $P(B|A)$  corresponde a probabilidade do evento  $B$  dado que o evento  $A$  tenha ocorrido e  $P(A|B)$  corresponde a probabilidade do evento  $A$  dado que o evento  $B$  tenha ocorrido.

A principal característica do algoritmo é que ele não leva em consideração a correlação entre as variáveis, tratando-as de forma independente. Essa característica é o que dá o nome de *naïve* (ingênuo) ao algoritmo.

## 4. Materiais e métodos

### 4.1 Ferramentas

Para a aplicação dos métodos para realizar a classificação de notícias foi utilizado a linguagem Python na versão 3.7.0, pela riqueza de bibliotecas de aprendizado de máquina e processamento de textos. A codificação foi realizada nas IDEs Visual Studio Code e Jupyter Notebook.

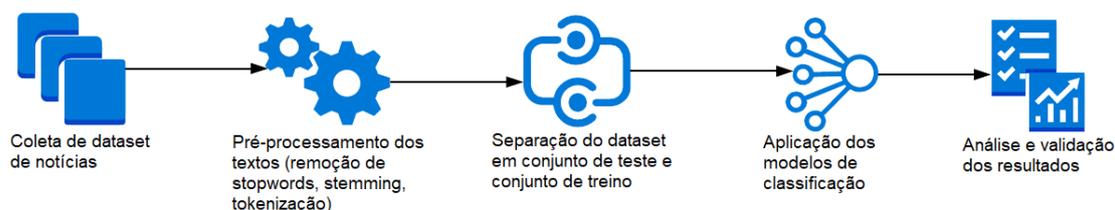
Quanto ao conjunto de dados para realizar as etapas de teste e treino dos algoritmos, foi utilizado o *dataset* de Monteiro (2018), que consiste num *dataset* de 5200 notícias, sendo que 2600 são verdadeiras e as outras 2600 são falsas. Os dados foram separados em duas pastas, uma para as notícias rotuladas como verdadeiras, outra para as notícias rotuladas como falsa, cada pasta contendo arquivos de texto (txt).

Além disso, é necessário a busca de um conjunto de *word embeddings* em português, para a aplicação do modelo proposto por Lai et al. (2015). O Núcleo Interinstitucional de Linguística Computacional da USP, produziu um repositório de *word embeddings* (Hartmann, 2017), com mais de 1 bilhão de tokens.

Para as bibliotecas, foram instalados pacotes como, TensorFlow para construção de redes neurais, nltk, spaCy e unidecode para leitura e processamento dos textos, scikit-learn para coletar métricas e implementação de um classificador probabilístico, matplotlib e seaborn para geração de gráficos.

## 4.2 Metodologia

O modelo do trabalho é classificador de textos, baseado numa representação atual como a de *word embeddings* Lai et al. (2015), aplicado sobre um *dataset* de notícias rotuladas, ou seja, que já foram previamente classificadas como sendo verdadeiras ou falsas. A figura 3 descreve as fases de um modelo de classificação.



**Figura 3. Diagrama com as fases do processo de classificação de notícias**

Fonte: O autor

É de extrema importância utilizar técnicas de processamento de linguagem natural para pré-processamento dos textos do conjunto de notícias. O texto bruto se apresenta de forma esparsa e desestruturada, que não permite a execução de classificadores de maneira eficiente. Vale destacar que a fase de pré-processamento de dados, pode ser considerada a fase mais importante que aplicação de qualquer tarefa de aprendizagem de máquina.

Métodos aplicados são (Bird et al., 2009):

- remoção de *stopwords*, que são palavras como artigos, verbos de ligação, que aparecem nos textos várias vezes, mas praticamente não influenciam a classificação;
- stemização ou lematização, que é a redução de palavras a seus radicais, removendo flexões de tempo verbal, gênero, número;
- tokenização que é o processo de criação de um vetor de termos de um documento, onde cada termo ocupa um índice do vetor

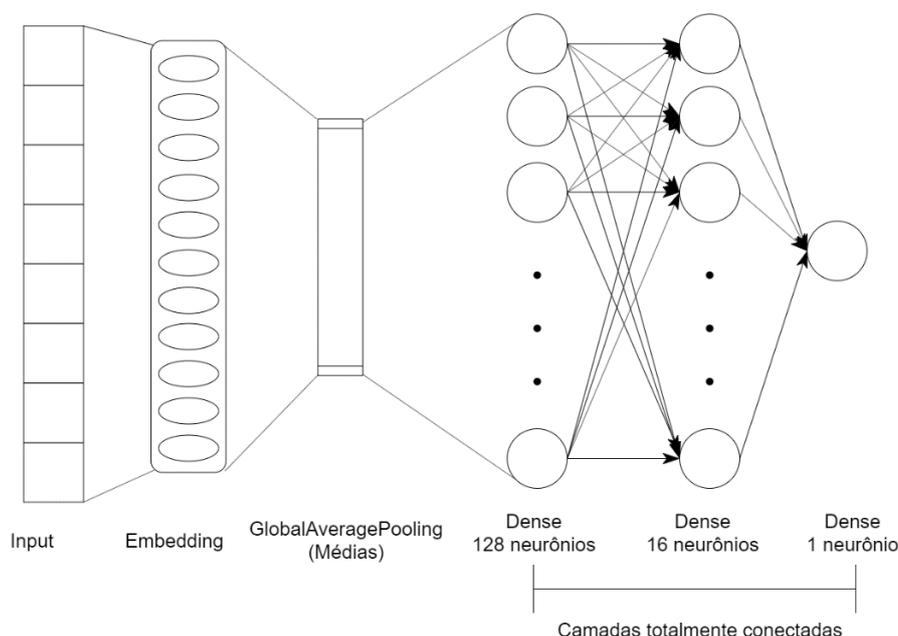
Após o pré-processamento, é necessário dividir o conjunto de notícias em dois: o conjunto de treino e o conjunto de teste, em uma proporção de 70% para o conjunto de treino e 30% para o de teste. O conjunto de treino é aquele no qual o algoritmo do modelo de aprendizado de máquina irá aprender a classificar as notícias. No caso das redes neurais, o treinamento consiste no ajuste de uma matriz de pesos, esta que será o “cérebro” da rede neural. Os pesos são valores decimais e esse ajuste consiste em aproximar o valor da saída, ou seja, a classe de notícia verdadeira ou falsa, com o valor desejado, que é a saída esperada desse documento que já foi classificado.

Nesse trabalho, a partir do *corpus* de notícias reunido por Monteiro (2018), realiza-se os pré-processamentos já citados. Em seguida, é montado um “dicionário”, no qual cada palavra é representada por um número inteiro, com as dez mil palavras mais frequentes do conjunto de notícias.

A escolha do método apresentado por Lai et al. (2015), teve como motivação testar o modelo de classificação proposto pelo autor e validar se a representação mais atual por meio de *word embeddings* com um classificador de uma rede de *deep learning* é válido, representando um ganho real, com um custo computacional razoável, quando comparado

com os modelos focados em uma representação clássica de matriz termo-documento, no domínio de identificação de *fake news*, na língua portuguesa, que é o foco do trabalho.

Na rede neural, temos uma estrutura com seis camadas: a primeira é a camada de entrada que recebe os documentos pré-processados, a segunda é de *embedding*, que é responsável por montar os vetores de cada palavra, a terceira é uma camada que realiza a média dos vetores de cada palavra, por fim as três camadas seguintes são camadas totalmente conectadas, que serão responsáveis pela classificação. Durante a fase de treinamento, o modelo irá aprender a classificar as notícias a partir do conjunto de testes, e o critério de parada do treinamento é atingir cinquenta épocas, ou quando o modelo parar de evoluir. A figura 4 apresenta o modelo descrito.



**Figura 4. Diagrama com a estrutura da rede neural**

Fonte: O autor

Para fins de comparação, também foi implementado um classificador que utiliza a representação de matriz termo-documento. Neste, as fases de pré-processamento do *corpus* continuam as mesmas, porém utilizando a matriz com termos e documentos, com as frequências dos termos de cada documento. O modelo em questão é o classificador probabilístico Naïve Bayes. A classe que tem a maior probabilidade é definida como a previsão.

Após o processo de aprendizado, o modelo precisa ser aplicado no conjunto de teste. Nele o modelo não possui a fase de aprendizado, no caso da rede neural o ajuste de pesos, e as notícias do conjunto de teste são diferentes do conjunto de treino.

#### 4.2.1 Métodos de avaliação de resultados

A partir da aplicação do modelo é necessário medir a precisão do modelo, taxa de acerto e erro, utilizando técnicas de estatística.

A taxa de acerto é uma métrica que informa a razão entre os valores previstos e os valores reais. A taxa de acerto é dada por:

$$\text{Taxa de acerto} = \frac{VP + VN}{VP + FP + VN + FN}$$

em que  $VP$  é o número de verdadeiros positivos,  $VN$  o número de verdadeiros negativos,  $FP$  o número de falsos positivos e  $FN$  o número de falsos negativos.

Há também a sensibilidade, também conhecida por taxa de verdadeiros positivos. Essa métrica mede a proporção dos valores previstos positivos com os verdadeiros positivos. A sensibilidade (TPR) é dada por:

$$TPR = \frac{VP}{VP + FN}$$

em que  $VP$  corresponde ao número de verdadeiros positivos e  $FN$  ao número de falsos negativos.

Temos também a especificidade, também conhecida por taxa de verdadeiros negativos. Essa métrica mede proporção dos valores previstos negativos com os verdadeiros negativos. A especificidade é dada por:

$$TNR = \frac{VN}{VN + FP}$$

em que  $VN$  é o número de verdadeiros negativos e  $FP$  é o número de falsos positivos.

O valor preditivo positivo (PPV) corresponde ao valor da proporção de verdadeiros positivos entre todas as ocorrências positivas. Esse valor mostra a probabilidade de uma ocorrência positiva ser realmente da classe positiva. É dado por:

$$PPV = \frac{VP}{VP + FP}$$

em que  $VP$  corresponde ao número de verdadeiros positivos e  $FP$  é o número de falsos positivos.

Já o valor preditivo negativo (NPV) corresponde ao valor da proporção de verdadeiros negativos entre todas as ocorrências negativas. Esse valor mostra a probabilidade de uma ocorrência negativa ser realmente da classe negativa. É dado por:

$$NPV = \frac{VN}{VN + FN}$$

em que  $VN$  é o número de verdadeiros negativos e  $FN$  ao número de falsos negativos.

A matriz de confusão é uma métrica que permite fazer uma análise visual que permite identificar a relação entre os valores previstos pelo classificador e os valores reais, assim como visualizar verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos (Powers, 2011). Ela apresenta os valores previstos e reais de forma através de forma tabular.

A curva ROC é uma métrica amplamente utilizada em modelos de classificação binária (de duas classes), pois é capaz de informar a qualidade do modelo em distinguir a classe positiva da classe negativa.

Segundo Narkhede (2018), ROC é uma curva de probabilidades e quanto maior a área abaixo da curva (AUC), melhor o modelo. Para criar essa curva são criados vários

limiares de separação entre as classes. O gráfico resultante é uma relação entre a taxa de verdadeiros positivos no eixo Y, e falsos positivos no eixo X.

O modelo de aprendizado foi validado por meio de validação cruzada (Refaeilzadeh, 2008), que consiste em rodar o processo de treinamento e teste  $n$  vezes, com diferentes amostras de treino e teste, geradas aleatoriamente a cada execução.

## 5. Resultados e discussões

Nos primeiros testes o modelo atingiu uma taxa de acerto entre 87% e 88%. A partir desses resultados, decidiu-se utilizar outro método para selecionar as dez mil palavras mais relevantes, o método do *chi-square*, ao invés da frequência. De acordo com Gajawada (2019), o método *chi-square* permite medir a dependência entre os atributos e as classificações, desse modo, usamos os dez mil melhores termos a partir desse método e então o modelo é treinado novamente.

Na figura 4 é apresentado a matriz de confusão do modelo de rede neural. A partir dela percebe-se que o modelo conseguiu diferenciar de maneira satisfatória as notícias verdadeiras das falsas, evidenciado pelo número elevado de verdadeiros positivos e verdadeiros negativos, e pelo baixo número de falsos positivos e falsos negativos.



**Figura 4: Matriz de confusão do teste do modelo de rede neural**  
Fonte: O autor

Após o uso do método *chi-square* para realizar a seleção de termos a precisão do modelo subiu para valores entre 91% e 93% de acerto, respectivamente para notícias falsas e verdadeiras, o que indica um resultado muito satisfatório.

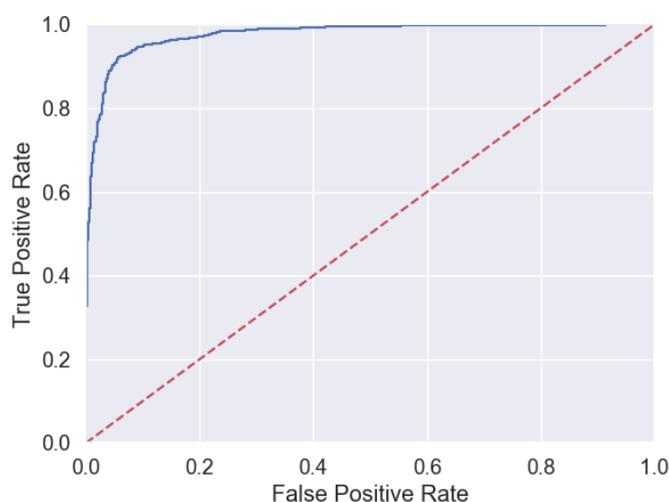
O quadro 1 mostra o resultado das métricas aplicadas no modelo de rede neural. A sensibilidade de 93% indica que o modelo teve um bom desempenho na classificação de textos da classe verdadeira. No modelo a especificidade atingiu 91%, que indica que o modelo teve um bom desempenho na classificação de textos da classe falsa.

Métrica	Classe verdadeira	Classe falsa
Precisão	91% (PPV)	93% (NPV)
Recall	93% (sensibilidade)	91% (especificidade)

**Quadro 1: Relação de métricas do modelo de rede neural**

A seguir foi aplicado o método de validação cruzada, aplicando o modelo dez vezes em diferentes amostras de conjunto de treino e teste. Como o modelo após cada treino pode mudar sua performance, mesmo que de maneira sensível, essa técnica permite fornecer um valor mais assertivo da taxa de acerto do modelo a partir da média dos k testes do modelo. O resultado da validação cruzada forneceu uma média de acurácia do modelo de 91%, com intervalo de confiança de 0,02.

A figura 5 mostra a curva ROC do modelo de rede neural. A partir da mesma é possível observar uma curva bem aberta, com uma grande área abaixo da curva (AUC). Com isso é possível dizer que o modelo realiza um bom trabalho para realizar a distinção entre as classes verdadeira e falsa em suas previsões.



**Figura 5: Curva ROC do modelo de rede neural**  
**Fonte: O autor**

Também foram realizados experimentos com o classificador probabilístico Naïve Bayes, que utiliza a representação de matriz de termo-documento ao invés de *word embeddings*. Nesse caso, a fase de pré-processamento foi o mesmo da rede neural, assim como as métricas aplicadas para medir a qualidade desse modelo.

A figura 6 mostra a matriz de confusão do modelo, percebe-se que apesar possuir um bom desempenho, a quantidade de verdadeiros positivos e verdadeiros negativos diminuíram, falsos positivos e falsos negativos aumentaram sensivelmente em comparação com o modelo de rede neural.



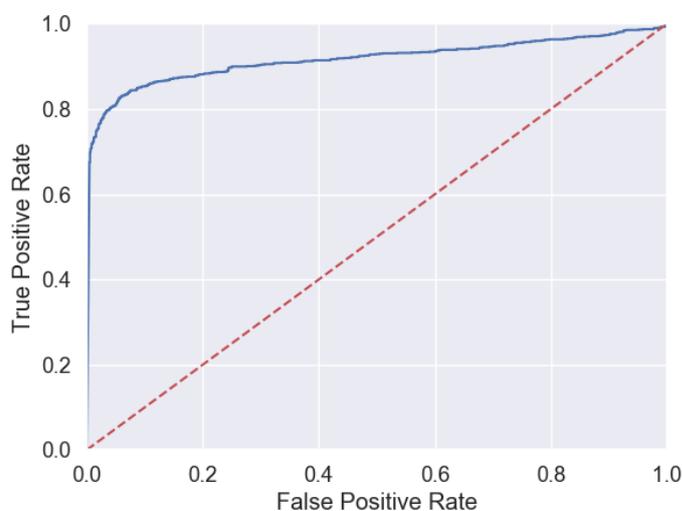
**Figura 6: Matriz de confusão do modelo Naïve Bayes**  
 Fonte: O autor

O quadro 2 permite observar que também houve uma leve diminuição nos valores de sensibilidade e especificidade do modelo, quando comparado com a rede neural. Isso mostra que o modelo de rede neural consegue diferenciar melhor as classes.

Métrica	Classe verdadeira	Classe falsa
Precisão	84% (PPV)	87% (NPV)
Recall	88% (sensibilidade)	83% (especificidade)

**Quadro 2: Relação de métricas do modelo Naïve Bayes**

A curva ROC do modelo Naïve Bayes, na figura 7, também põe em evidência essa leve perda de precisão do modelo em relação ao de rede neural, ao apresentar uma curva um pouco mais fechada e uma área abaixo da curva menor.



**Figura 7: Curva ROC do modelo Naïve Bayes**  
 Fonte: O autor

## 5. Conclusão

Nesse trabalho, percebeu-se como a evolução tecnológica teve o contraponto de aumento da disseminação de desinformação e notícias falsas, o impacto negativo que é causado e a dificuldade de combater por meio de métodos tradicionais.

Ficou evidente que o problema abordado nesse trabalho se trata de algo que vem sido amplamente abordado na literatura quanto ao processamento de linguagem natural, que é a classificação de textos. Na seção 2, foram vistos alguns trabalhos correlatos, não só restritos ao domínio do trabalho de classificação de notícias falsas, que permitiram ter uma perspectiva da diversidade dos métodos de abordagem para resolução do problema, como também quanto a maneira representar os dados.

Como contribuição, os experimentos conduzidos neste trabalho trouxeram uma solução alternativa em relação ao que foi apresentado por Monteiro (2018), que também aborda o domínio de *fake news*, no qual o autor deu preferência ao uso de SVM como classificador, e a rede neural nesse trabalho apresentou ganho em acurácia, ainda que pequeno, em bom tempo computacional. Além disso, houve o uso do método *chi-square* para seleção de termos mais relevantes, que não foi utilizado em alguns trabalhos correlatos da seção 2, principalmente os que envolvem processamento de textos em português.

Desse modo, a solução proposta nesse trabalho mostra ser promissora no auxílio de combate a notícias falsas. Os resultados dos experimentos conduzidos, mostram que o modelo baseado em rede neural, com uma representação mais atual de *word embeddings*, possui resultados superiores do que um modelo baseado numa representação em matriz termo-documento. Apesar disso, não se deve descartar totalmente modelos baseados nessa representação mais clássica de textos, porque também apresentou bons resultados, e dependendo do problema que precisa ser tratado essa abordagem pode ser mais adequada.

Destaca-se que para que esse modelo seja realmente efetivo para auxiliar a combater notícias falsas, é necessário que eventualmente ele seja retreinado. O motivo é que o modo de escrever essas notícias pode variar com o tempo, desse modo utilizar o modelo sem eventual retreinamento, pode deixá-lo com o tempo obsoleto e causar perda de eficiência.

Como trabalhos futuros, pretende-se buscar mais textos para realizar a comprovação da eficácia do modelo fora de um ambiente controlado, como no caso em que foram conduzidos os experimentos no trabalho. Além disso, realizar melhorias quanto o método de seleção de palavras de um documento, para dar maior assertividade que as palavras relevantes serão mantidas. Também se deseja explorar outras estruturas de rede neural, como recorrentes e convolucionais, em conjunto de um tempo computacional adequado, mesmo nesses modelos mais robustos.

## 6. Referências

Ahire, J. B. “The Artificial Neural Networks Handbook” DZone. Disponível em: <<https://dzone.com/articles/the-artificial-neural-networks-handbook-part-1-1>> Acesso em 23 out 2019

Ahmed, H., Traore, I., Saad, S. “Detecting opinion spams and fake news using text classification”. ECE Department, University of Victoria, Victoria, Canadá, 2017

Becker, L. “Algoritmo de Classificação Naive Bayes.” Disponível em <<https://www.organicadigital.com/blog/algoritmo-de-classificacao-naive-bayes/>> Acesso em 14 out 2019

Bird, S., Klein E., Loper E. “Natural Language Processing with Python” 1º ed. O’Reilly Media Inc, 2009

Folha de S. Paulo. “Fake news' alteram hábitos do público, indica pesquisa.” Disponível em:<<https://www1.folha.uol.com.br/mundo/2017/10/1931635-fake-news-alteram-habitos-do-publico-indica-pesquisa.shtml>> Acesso em 04 mar 2019

G1. “EUA fazem o maior estudo sobre fake news nas redes sociais.” Disponível em: <<http://g1.globo.com/jornal-nacional/noticia/2018/03/eua-fazem-o-maior-estudo-sobre-fake-news-nas-redes-sociais.html>> Acesso em 04 mar 2019

Gajawada, S. K. “Chi-Square Test for Feature Selection in Machine learning” Disponível em: < <https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223>> Acesso em 16 out 2019

Hartmann, N. S. et al. “Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks”. In: Symposium in Information and Human Language Technology, 11., 2017, Uberlândia. p. 122-131

Hunt, Elle. “What is fake news? How to spot it and what you can do to stop it.” The Guardian. Disponível em: <<https://www.theguardian.com/media/2016/dec/18/what-is-fake-news-pizzagate>> Acesso em 03 mar 2019

Karani, D. “Introduction to Word Embedding and Word2Vec” Disponível em <<https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>> Acesso em 14 out. 2019

Lai, S. et al. “Recurrent Convolutional Neural Networks for Text Classification” In: Twenty-Ninth AAAI Conference on Artificial Intelligence

LeCun, Y. “Deep Learning”, Nature, v. 521, p. 436 – 444, maio 2015

Monteiro R.A., Santos R.L.S., Pardo T.A.S., de Almeida T.A., Ruiz E.E.S., Vale O.A. (2018) “Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results.” In: Villavicencio A. et al. (eds) Computational Processing of the Portuguese Language. PROPOR 2018. Lecture Notes in Computer Science, vol 11122. Springer, Cham

Narkhede S. “Understanding AUC - ROC Curve” Disponível em: < <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>> Acesso em 27 set. 2019

Powers, David M. W. “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation”. School of Informatics and Engineering, Flinders University of South Australia, 2007

Refaeilzadeh, P., Tang, L., Liu, H. “Cross-Validation”. Arizona State University, 2008

Ross, S. “A First Course in Probability” 8º ed. Pearson Prentice Hall, 2010

Rossi, R. G. “Classificação automática de textos por meio de aprendizado de máquina baseado em redes”. Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), 2015

Saha, S. “A Comprehensive Guide to Convolutional Neural Networks”, Towards Data Science. Disponível em < <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>> Acesso em 23 out 2019

Santos, F. “Mineração de opinião em textos opinativos utilizando algoritmos de classificação”. Departamento de Ciência da Computação, Universidade de Brasília (UnB), 2013

Vaz, A. L. “Introdução teórica a Neural Network – Deep Learning” Disponível em <<https://medium.com/data-hackers/neural-network-deep-learning-parte-1-introdu%C3%A7%C3%A3o-te%C3%B3rica-5c6dcd2e5a79>> Acesso em 23 de out 2019

Zannata, Rafael A. F. “Eleições e Fake News: o tortuoso caminho do Brasil.” Outras Palavras. Disponível em: <<https://outraspalavras.net/sem-categoria/eleicoes-e-fake-news-o-tortuoso-caminho-do-brasil/>> Acesso em 04 mar 2019