

FACULDADE DE ENGENHARIA DE SOROCABA - FACENS
PÓS GRADUAÇÃO EM ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

ALEX COELHO ABRANTES

BRUNO ALVES COMITRE

DETECÇÃO AUTOMATIZADA DE NOTÍCIAS FALSAS:
PESQUISA COM RECONHECIMENTO DE INTEGRIDADE DAS INFORMAÇÕES

Tese apresentada ao Programa de Pós-Graduação
em Ciências de Dados da Faculdade de Engenharia de
Sorocaba - FACENS, como requisito parcial para a obtenção
do título de Pós Graduado em Ciência de Dados.

Orientado: Prof. Matheus Mota

Coord.: Prof. Fernando Vieira da Silva

SOROCABA

2019

DEDICATÓRIA

Dedicamos esta tese a Alan Turing, que proporcionou o campo da ciência da computação, onde podemos ter toda evolução tecnológica mencionada de forma que foi possível a criação do que temos como evolução tecnológica hoje e que podemos sonhar com todos os avanços tecnológicos ao qual seres humanos consegue criar; e aos nossos pais, que a despeito das nossas muitas ausências, sempre estiveram ao nosso lado.

AGRADECIMENTOS

À Universidade gostaríamos de deixar uma palavra de gratidão por ter recebido de braços abertos e com todas as condições que nos proporcionaram dias de aprendizagem muito ricos, durante todos os períodos.

Aos professores reconhecemos o esforço gigante com muita paciência e sabedoria. Foram eles que nos deram recursos e ferramentas para evoluir um pouco mais todos os dias.

É claro que não podemos esquecer família e amigos, porque foram eles que incentivaram e inspiraram através de gestos e palavras a superar todas as dificuldades e ausências.

A todas as pessoas que alguma forma nos ajudaram a acreditar que o esforço valia a pena, queremos deixar um agradecimento.

RESUMO

Abrir o Google, digitar “fake news”, pesquisar, encontraremos diversas definições, mas, resumidamente “fake news” que em português significa “Notícia Falsa”, é a forma de propagação de desinformações por meio de comunicação com a intenção de enganar a fim de obter algum tipo de vantagem. São abordadas principalmente no contexto político, científico, social e valores cívicos, geralmente atraindo grupos de pessoas que buscam a princípio a atenção por informações rápidas e curtas, ou seja, notícias prontas sobre determinados assuntos que explana toda informação e opinião.

Além do alto impacto da propagação, existe a questão dos robôs de internet, são aplicações autônomas que rodam na internet executando algum tipo de tarefa pré-determinada. Criminosos e pessoas mal-intencionadas utilizam esses robôs para propagar conteúdos falsos pela internet, e foi muito utilizado nas campanhas eleitorais norte-americanas. Atualmente existe uma preocupação global no combate dessas “fábricas de mentiras”. A questão agora está em como intervir para diminuir esse tipo de propagação? Como identificar uma notícia falsa? Como impedir uma disseminação viral de notícias?

Além do contexto educacional, temos como abordagem deste trabalho a detecção e intervenção com base em algoritmo de aprendizado de máquina identificar textos caracterizados como uma notícia falsa.

PALAVRAS-CHAVE: *Fake News*. Dados. Comunicação. Aprendizado de Máquina.

ABSTRACT

Opening Google, typing “fake news”, searching, we will find several definitions, but in short “fake news” which in Portuguese means “False News”, is the form of spreading misinformation through communication intended to mislead in order to to gain some kind of advantage. They are mainly addressed in the political, scientific, social and civic context, generally attracting groups of people who seek attention at first for quick and short information, ie ready-made news on certain subjects that explain all information and opinion.

Besides the high impact of propagation, there is the issue of internet robots, they are standalone applications that run on the internet performing some kind of predetermined task. Criminals and malicious people use these robots to spread fake content over the Internet, and it has been widely used in US election campaigns. There is currently a global concern in fighting these “lie factories”. The question now is how to intervene to slow this kind of spread? How to identify fake news? How to prevent a viral spread of news?

In addition to the educational context, we have as approach to this work the detection and intervention based on machine learning algorithm identify texts characterized as false news.

KEYWORDS: *Fake News*. Data. Communication. Machine Learning.

SUMÁRIO

- [1. FAKE NEWS](#)
 - [2. REDES NEURAIAS](#)
 - [2.1 Redes Neurais Recorrentes](#)
 - [2.2 Métodos de Aprendizado](#)
 - [2.3 Long Short-term Memory](#)
 - [2.4 Padrões de Texto com Python](#)
 - [3. ANÁLISE EXPLORATÓRIA DOS DADOS](#)
 - [3.1 Dados Faltantes](#)
 - [3.2 Novos Recursos \(Features\)](#)
 - [3.3 Dados Desequilibrados e Randon Shuffle](#)
 - [4. ANÁLISE GRÁFICA DOS DADOS](#)
 - [4.1 Word Cloud](#)
 - [4.2 Histograma](#)
 - [4.3 Boxplot](#)
 - [4.4 Dispersão](#)
 - [4.5 Correlação](#)
 - [5. IMPLEMENTAÇÃO](#)
 - [5.1 Modelo em Keras TensorFlow](#)
 - [6. RESULTADO FINAL](#)
 - [REFERÊNCIAS](#)
-

1 FAKE NEWS

Réflexions d'Un Historien Sur les Fausses Nouvelles de la Guerre, que em português significa ("Reflexões de um historiador sobre as falsas notícias da guerra", Allia, 2012) é o título de um texto publicado originalmente em 1921 por Marc Bloch, um dos mais influentes historiadores do século XX, após retornar da Primeira Guerra Mundial alucinado com a importância que as notícias falsas haviam tido, ao refletir sobre escreveu: "Histórias falsas aumentaram a multidão. As notícias falsas, em toda a multiplicidade de suas formas - contos simples, imposturas, lendas - encheram a vida da humanidade. ", continuou, "Como elas nascem? De quais elementos elas extraem? Como elas se espalham, ganhando impulso quando passam de boca em boca ou escrevendo por escrito? Não há dúvida mais do que aqueles que merecem fascinar quem gosta de pensar em história. " E afirmou, "Um erro só se propaga e se amplifica, só ganha vida com uma condição: encontrar um caldo de cultivo favorável na sociedade onde se expande. Nele, de forma inconsciente, os homens expressam seus preconceitos, seus ódios, seus temores, todas as suas emoções". Em outras palavras, as notícias falsas necessitam de pessoas que preferem informações que confirmam suas atitudes preexistentes, veem as informações consistentes com suas crenças como mais persuasivas do que as dissonantes e tendem a aceitar informações que as agradam.

Um estudo que avaliou a divulgação de importantes notícias falsas estimou que o americano encontrou entre uma a três notícias falsas durante o mês anterior às eleições americanas de 2016. Outro estudo relatou que informações falsas no Twitter são normalmente retuitadas por muito mais pessoas, e muito mais rapidamente, do que informações verdadeiras, especialmente quando o assunto é política, as pessoas têm muito mais probabilidade de acreditar em histórias que favorecem seu candidato preferido. Pesquisadores das universidades do Sul da Califórnia e de Indiana estimaram que haja entre 9% a 15% de robôs no Twitter e o Facebook estimou que 60 milhões de robôs podem estar infestando sua plataforma. Eles foram responsáveis por uma parte substancial do conteúdo político publicado durante a campanha nos EUA de 2016, e alguns dos mesmos robôs foram usados mais tarde para tentar influenciar a eleição francesa de 2017. Portanto, a identificação dessas notícias falsas e a tentativa de impedir que não se tornem virais nas mídias sociais, é um grande desafio de pesquisa.

O desafio proposto é criar um modelo em aprendizado de máquina para reconhecimento de linguagem humana ou natural sobre um conjunto de dados que contenha textos, e seja capaz de identificar se o texto se refere a notícia verdadeira ou se trata de uma notícia falsa, as chamadas **Fake News**.

No início dos estudos nosso desafio era extrair tais informações por meio de sites jornalísticos e blogs, mas, entendemos que a demanda de tempo seria grande e o foco do trabalho é o desenvolvendo de um modelo para aprendizado de máquina que seja capaz de identificar uma notícia falsa, então a tarefa de scraping foi cancelada e procuramos encontrar algum conjunto de dados que atendesse a estrutura que havíamos planejado:

```
{
  "article": {
    "id": "number sequencial",
    "type": "object",
    "url": "string",
    "base_url": "string",
    "main_language": "string",
    "authors": "string",
    "title": "string",
    "subtitle": "string",
    "body": {
      "paragraphs": "qty paragraphs",
      "paragraphs_body": "array each paragraphs",
      "paragraphs_main_entities": "array each main entities",
      "number_of_words": "string",
      "tags": "string",
      "images_url": "string",
      "image_legends": "string",
      "datetime": "string"
    }
  }
}
```

Segundo Victoria L. Rubin et al. 9 pontos são importantes para a coleta dos dados:

- - Paridade de notícias verdadeiras e falsas
 - Notícias em versão texto
 - Possibilidade de confirmar a verdade
 - Homogeneidade no tamanho dos artigos
 - Homogeneidade no estilo de escrita
 - Janela de tempo bem definida para coleta
 - Definir o estilo de notícia
 - Fatores pragmáticos, como custo de aquisição e disponibilidade das notícias
 - Definir idioma e cultura das notícias

Encontramos um conjunto de dados em um desafio do site Kaggle, intitulado: Fake News - Build a system to identify unreliable news articles. A pesquisa contém 20.000 dados extraídos na web e estão rotulados como confiável e não confiável. O conjunto de dados é composto por artigos de âmbito político e no idioma inglês, portanto todo o estudo será com base neste idioma.

- **Link:** <https://www.kaggle.com/c/fake-news/data> (<https://www.kaggle.com/c/fake-news/data>)

</p>

O Schema do conjunto de dados se aproximou com o que estávamos planejando, e por esse motivo nos

2 REDES NEURAIS

Biologicamente o cérebro humano ao qual possui bilhões de neurônios conectados entre si, constituído por um corpo celular ligado por ramificações, chamadas de dendritos, onde conhecidos como axônio, possuem um terminal denominado de telodendro. E esta sinapse que faz a ligação entre dois neurônios, ligados e através de pulsos nervosos se comunicam. Estes impulsos recebidos são processados até atingir um limiar de ação, que produz uma substância neurotransmissora que flui do corpo celular para o axônio que está conectada a um dendrito de outro neurônio. Inspirado pelo sistema nervoso central a ciência da computação desenvolveu técnicas computacionais que apresentam modelos matemáticos capazes de adquirirem conhecimentos através de experiências e padrões.

O reconhecimento de padrões funciona como um algoritmo de aprendizado de máquina que ajustam a saída para que os neurônios ou nós interconectados entre si, formam uma rede complexa de processamento ao qual essas informações a fim de serem utilizadas para aplicações em controle de processos em diversas áreas, desde o reconhecimento de fala, quanto escrita e padrões.

2.1 Redes Neurais Recorrentes

Através dos estudos das Redes Neurais biológicas caracterizaram-se as Redes Neurais Recorrentes (RNN). Perceberam que com esta metodologia poderia resolver problemas de finalidades não algorítmicas ao simular reconhecimento de resolução de problemas que apenas o cérebro humano seria capaz de realizar. Com isso criou-se o conceito das RNNs. Geralmente esta rede é conectada por canais de comunicação associados em operações de entrada e saída entre as conexões, gerando interações entre os nós desta rede a fim de criar um comportamento inteligente. As Redes Neurais Recorrentes têm como características sua capacidade de fazer armazenamento associativo.

O objetivo do campo é criar modelos que se aproximem do funcionamento de sistemas biológicos dos seres humanos. E para isso são realizados modelos de redes neurais voltados a teoria biológica para o aprendizado estatístico e teoria da informação.

2.2 Métodos de Aprendizado

Existem inúmeros tipos de algoritmo de aprendizado de máquinas e ao estudo de RNNs. Estes algoritmos se diferenciam entre si pelo método de classificação e modificação dos pesos de suas conexões. Entre os métodos mais conhecidos podemos ter como referências:

- Aprendizado Supervisionado; - Aprendizado Não Supervisionado; - Aprendizado Hebbiano; - Aprendizado por Reforço; - Aprendizado por Competição.

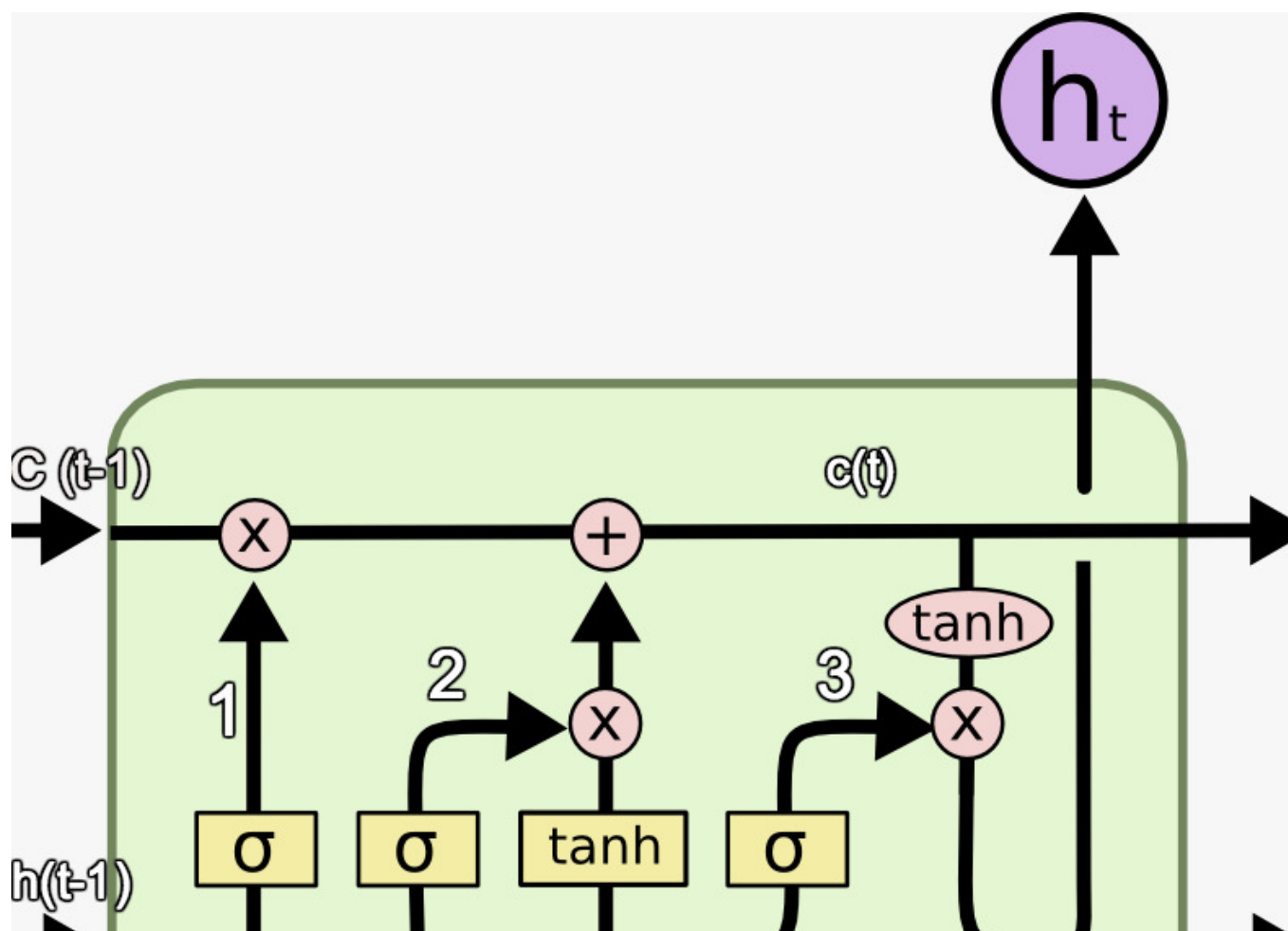
O método utilizado nesta pesquisa teve como base o aprendizado de máquina supervisionado e por reforço.

2.3 Long Short-term Memory

A sigla LSTM, em inglês (Long Short Term Memory), é a arquitetura de rede neural recorrente de memória longa, ou seja, em determinado intervalo arbitrário de valores, é guardado valores antigos, ao qual a rede "lembra" destes valores. Sendo assim a rede recorrente de longo prazo é adequada para classificar, processar e prever séries temporais com intervalos de tempo de duração desconhecida.

Assim sendo uma vantagem da LSTM em relação a RNNs tradicionais. Uma arquitetura comum em LSTM é composta por uma célula (a parte da memória da unidade LSTM) e três "reguladores", do fluxo de informações dentro da unidade LSTM, composta ainda por uma porta de entrada, uma porta de saída e uma porta de esquecimento.

Existem variações, este não é um modelo exclusivo de uso, possuindo assim mais portões ou mais saídas. Algumas das famosas aplicações das LSTMs incluem, Modelagem de Linguagem, Tradução de Idiomas, Legendas em Imagens, Geração de Texto, e Chatbots.



2.4 Padrões de Texto com Python

Este trabalho utilizou a linguagem de programação Python. Assim como nos gráficos, insights e no modelo de aprendizagem de máquina para o reconhecimento de padrões de texto e caracteres utilizados a fim identificar de forma mais abstratos conceitos linguísticos com a abordagem das RNNs para separar orações completas e através de neurônios de saída e assim classificar seus significados.

3 ANÁLISE EXPLORATÓRIA DOS DADOS

Inicialmente, foi aplicado de técnicas para manusear valores faltantes e fazer transformações de variáveis. Os dados serão ajustados e estreitando os pressupostos para empregar técnicas gráficas e quantitativas, visando maximizar a obtenção de informações, tendências e detecção de comportamentos.

Ao importar os dados, é importante entender e identificar o intervalo de preditores específicos, identificar o tipo de dados de cada preditor, bem como calcular o número ou a porcentagem de valores omissos para cada preditor. Foi utilizado a biblioteca `pandas_profiling`, esta biblioteca fornece muitas funções extremamente úteis para a análise exploratória de dados.

Overview

Dataset info

Number of variables	5
Number of observations	20800
Missing cells	2554 (2.5%)
Duplicate rows	0 (0.0%)
Total size in memory	812.6 KiB
Average record size in memory	40.0 B

Variables types

Numeric	1
Categorical	3
Boolean	1
Date	0
URL	0
Text (Unique)	0
Rejected	0
Unsupported	0

Warnings

- `author` has a high cardinality: 4202 distinct values
- `author` has 1957 (9.4%) missing values
- `text` has a high cardinality: 20387 distinct values
- `title` has a high cardinality: 19804 distinct values
- `title` has 558 (2.7%) missing values

Warning

Missing

Warning

Warning

Missing

Variables

author
Categorical

Distinct count	4202
Unique (%)	20.2%
Missing (%)	9.4%
Missing (n)	1957

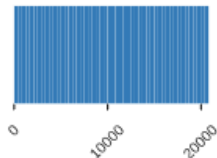
Pam Key	243
admin	193
Jerome Hudson	166
Other values (4198)	18241
(Missing)	1957

[Toggle details](#)

id
Numeric

Distinct count	20800
Unique (%)	100.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0

Mean	10399.5
Minimum	0
Maximum	20799
Zeros (%)	< 0.1%



[Toggle details](#)

label
Boolean

Distinct count	2
Unique (%)	< 0.1%
Missing (%)	0.0%
Missing (n)	0

1	10413
0	10387

[Toggle details](#)

text
Categorical

Distinct count	20387
Unique (%)	98.0%
Missing (%)	0.2%
Missing (n)	39

	75
Leave a Reply Click here to ge...	28
Click Here To Learn More Abo...	26
Other values (20383)	20632
(Missing)	39

[Toggle details](#)

title
Categorical

Distinct count	19804
Unique (%)	95.2%
Missing (%)	2.7%
Missing (n)	558

The Dark Agenda Behind Glob...	5
Get Ready For Civil Unrest: S...	5
Televisión: lo más visto ayer	4
Other values (19800)	20228
(Missing)	558

[Toggle details](#)

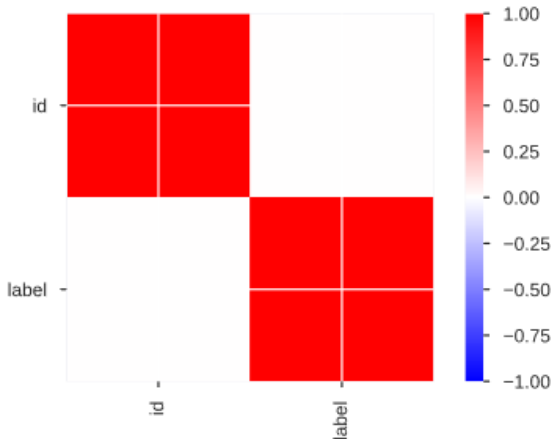
Correlations

Pearson's r

Spearman's ρ

Kendall's τ

Phik (ϕ_k)



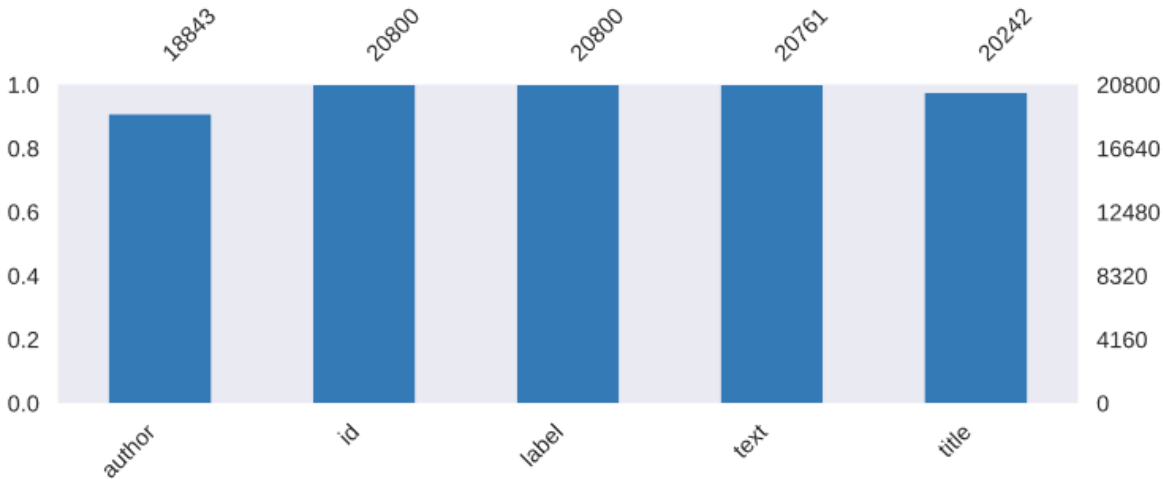
Missing values

Count

Matrix

Heatmap

Dendrogram



3.1 Dados Faltantes

A falta de dados pode afetar a análise e o treinamento, que poderá levar a falhas no aprendizado. É possível explorar se há dados ausentes no conjunto de dados através do relatório gerado do pandas_profiling. Observando o relatório identificou:

O título do atributo tem 558 amostras (2,68%) com valores ausentes.

O autor do atributo possui 1957 amostras (9,41%) com valores ausentes.

O texto do atributo tem 39 amostras (0,19%) com valores ausentes.

Como existe dados faltantes nos 3 recursos do dataset (Title, Author e Text) eliminar as linhas em que há dados ausentes neste caso é a melhor opção para não comprometer a análise e o treinamento.

O resultado da eliminação de 20.800 para 18.285 registros.

3.2 Novos Recursos (Features)

Após o tratamento dos dados, identificou-se a necessidade de criar novos recursos (features). A Análise de texto não é uma das tarefas mais fácil a se realizar quando os textos são muito extensos, embora, seja possível através de extração de informações por palavras obter uma melhor compreensão do corpus do texto. Podemos então criar novos recursos para estudar o comportamento do texto a fim de extrair informações como quantidade de palavras, comprimento do artigo e identificar as classes gramaticais como: substantivos, adjetivos e verbos.

Esses novos recursos serão utilizados nas análises gráficas procurando encontrar algum fenômeno ou padrão.

```
Index(['id', 'title', 'author', 'text', 'label', 'title_token',
      'title_token_distint', 'title_comprimento', 'title_num_palavras',
      'title_num_palavras_unicas', 'title_palavras_vs_unico',
      'title_substantivos', 'title_adjetivos', 'title_verbos',
      'title_contagem_substantivos', 'title_contagem_adjetivos',
      'title_contagem_verbos', 'text_token', 'text_token_distint',
      'text_comprimento', 'text_num_palavras', 'text_num_palavras_unicas',
      'text_palavras_vs_unico', 'text_substantivos', 'text_adjetivos',
      'text_verbos', 'text_contagem_substantivos', 'text_contagem_adjetivos',
      'text_contagem_verbos'],
      dtype='object')
```

3.3 Dados Desequilibrados e Random Shuffle

Continuando o tratamento nos dados após a eliminação dos registros faltantes, aplicou-se uma análise descritiva para verificar se houve desequilíbrio dos 18.285 registros. Analisando o valor da média do recurso "Label", responsável por armazenar os targets, 0 (zero) representa os artigos classificados como não confiáveis e 1 (um) representa os confiáveis, resultou em 43%, isso significa que os dados estão em desequilíbrio. Verificando a diferença têm 2.437 registros a mais para dados confiáveis. Para manter o conjunto de dados equilibrado aplicou-se a técnica de Random Shuffle nos registros com valor 1 assumindo o valor de 7.924 o número total de dados não confiáveis. Agora o conjunto de dados contém total de 15.848 registros distribuídos igualmente em 7.924 confiáveis e 7.924 não confiáveis.

```
count    18285.000000
mean      0.433361
std       0.495553
min       0.000000
25%       0.000000
50%       0.000000
75%       1.000000
max       1.000000
Name: label, dtype: float64
```

Não confiável: 7924

Confiável: 10361

Desequilíbrio nos dados de 2437 a mais para registros considerados confiáveis

Random Shuffle aplicado no conjunto

Não confiável: 7924

Confiável: 7924

4 ANÁLISE GRÁFICA DOS DADOS

Os gráficos são recursos importante para identificação de fenômeno que possa ser mensurado, quantificado ou ilustrado, fornecendo uma dimensão estatística sobre um determinado fato. Existem diversos tipos de gráficos e interpretar corretamente é de grande importância para compreender determinados fenômenos. Neste trabalho abordaremos apenas alguns tipos para responder a questões levantadas:

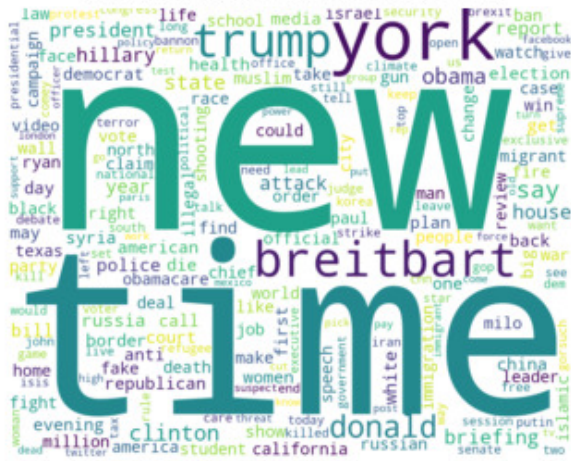
- Gerar um WordCloud com todos os títulos e textos e analisar as palavras mais utilizadas.
- Gerar um histograma referente ao tamanho dos títulos e textos. Será que existe diferença de tamanho (caracteres) para os títulos e textos confiáveis e não confiáveis?
- Gerar um boxplot referente a quantidade de palavras dos títulos e textos e analisar os valores mínimos, máximos, primeiro e terceiro quartil, mediana e existência de outliers. Será que existe diferença de entre confiáveis e não confiáveis?
- Podemos verificar alguma correlação entre os novos recursos?
- A forma de escrita do texto (exemplos: educado, rude, gírias, etc...) tem influência no sentimento Confiável e Não Confiável?

4.1 Word Cloud

A ideia central ao plotar o gráfico de nuvem de palavras, é encontrar o volume de vezes que determinadas palavra se repetem dentro do corpus do texto. Este tipo de gráfico se trata de um método heurístico a fim de encontrar respostas viáveis, ainda que imperfeitas, pois mesmo observando as palavras de destaques não há muito resultado para nosso problema, porém nos guiam para alguns questionamentos.

Plotamos 10 gráficos para compreender algum padrão entre os textos. Notamos que as palavras mais relacionadas quando se trata de fake news são ligadas a pessoa. Na observação notamos a frequência de Hillary Clinton e Trump. Não é por menos, pesquisas relatam que nas eleições norte americanas de 2016 muito se especulou o uso de bots e grupos disseminando notícias falsas atacando cada candidato.

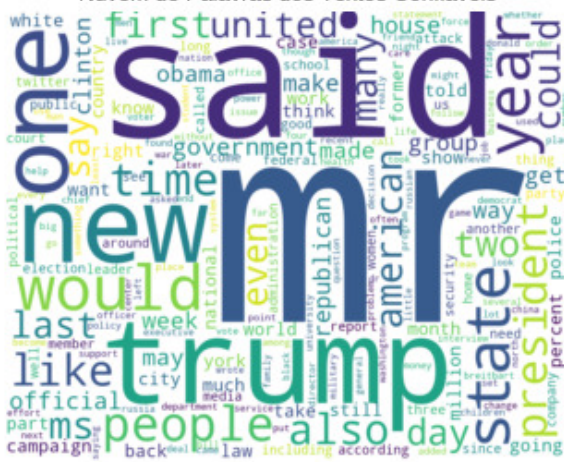
Nuvem de Palavras dos Títulos Confiáveis



Nuvem de Palavras dos Títulos Não Confiáveis



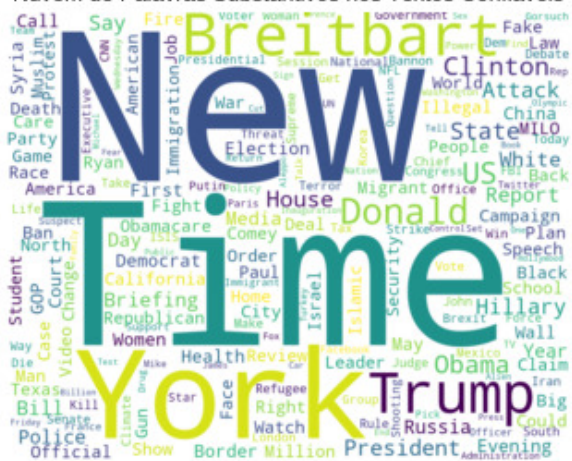
Nuvem de Palavras dos Textos Confiáveis



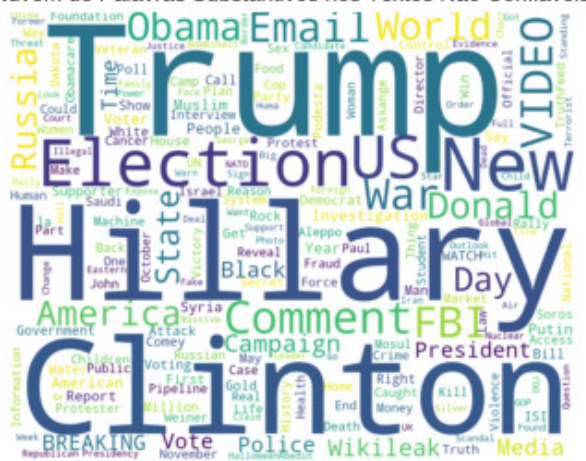
Nuvem de Palavras dos Textos Não Confiáveis



Nuvem de Palavras Substantivos nos Textos Confiáveis



Nuvem de Palavras Substantivos nos Textos Não Confiáveis



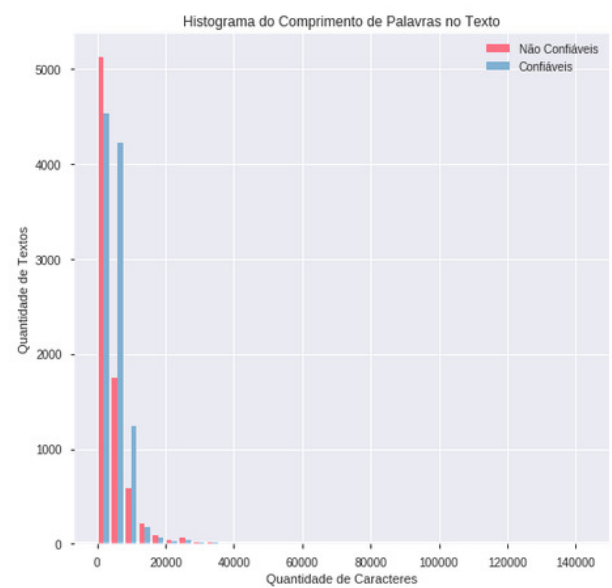
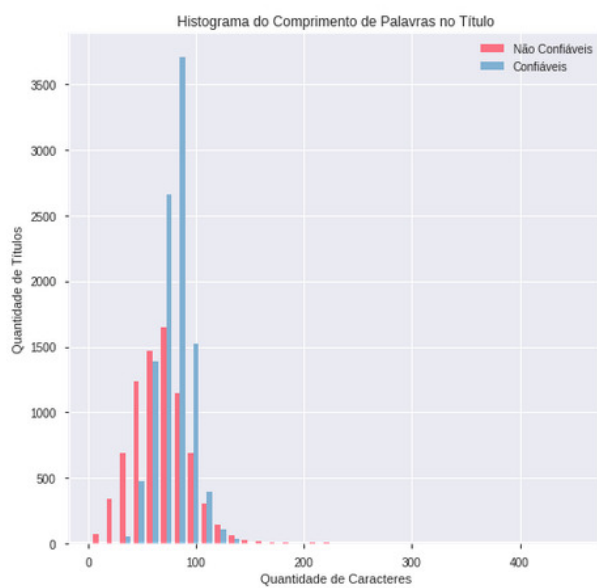
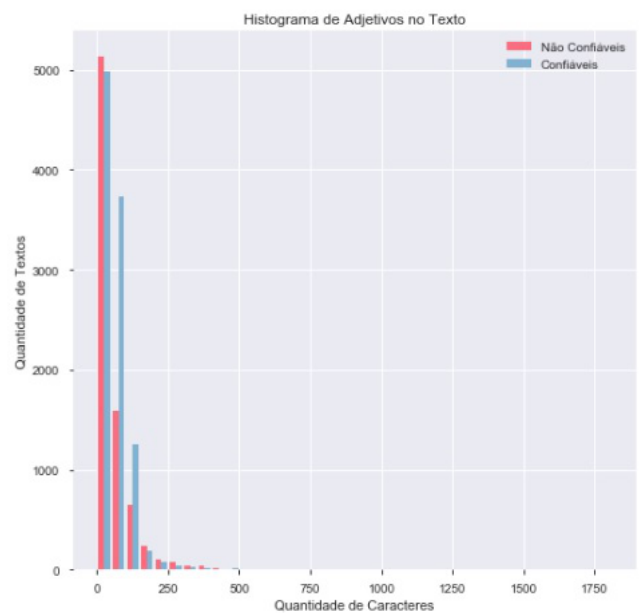
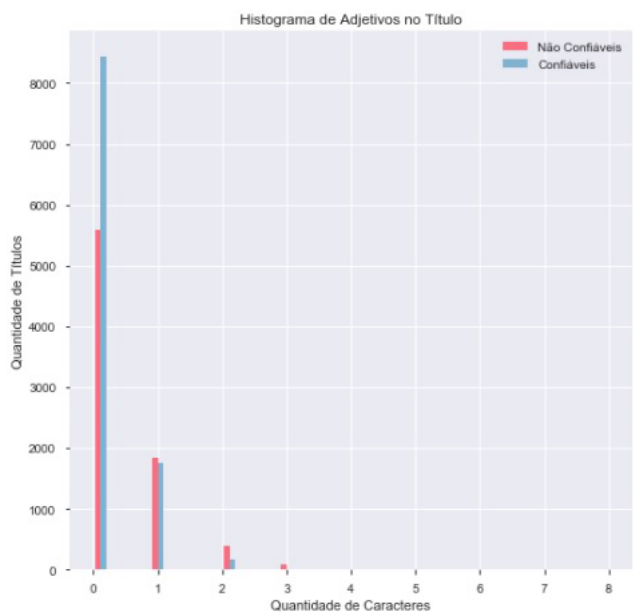
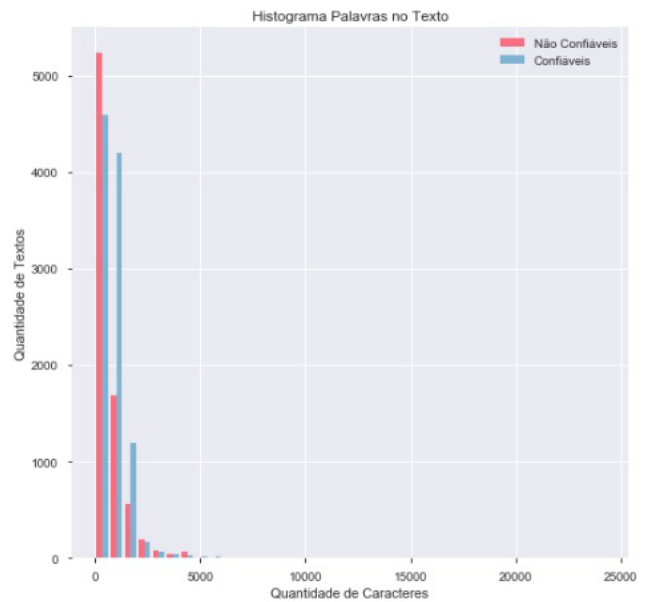
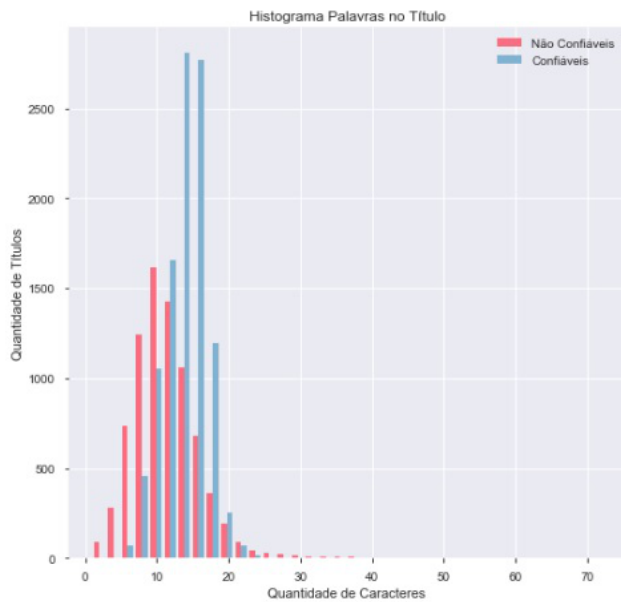
[illegible][illegible][illegible][illegible]

4.2 Histograma

A construção de histogramas tem caráter preliminar em qualquer estudo e é um importante indicador da distribuição de dados. Neste estudo utilizamos a frequência absoluta, que é o número que representa a quantidade de dados em uma determinada amostra ou o intervalo de classe específico, indicando a frequência (absoluta) com que uma classe aparece no conjunto de dados.

A seguir, os gráficos plotados abaixo representam os histogramas da frequência de palavras e adjetivos encontrados nos títulos e textos. Estes gráficos facilitam no entendimento e na análise da existência de influência na avaliação: Confiável ou Não Confiável.

Quando exploramos a frequência de palavras encontramos uma distribuição normal nos títulos que foram classificados como falsos, uma diferença em comparação aos títulos rotulados como confiáveis. Notamos que quando se trata de títulos confiáveis, tem tendência de ser maiores que 10 palavras e maiores que 50 caracteres, no qual os títulos mais curtos, são propícios a uma notícia falsa. Há uma evidência que notícias verdadeiras utilizam muito mais adjetivos em seus textos.



4.3 Boxplot

O diagrama de caixa é construído utilizando as referências de valores mínimos e máximos, primeiro e terceiro quartil, mediana e os outliers da base de dados. Diferentemente do histograma que possibilita ter uma melhor visualização das médias e desvio padrão, o diagrama de caixa têm como objetivo estudar as medidas estatísticas e identificar os valores atípicos dentro do conjunto de dados.

Com os gráficos de caixa plotados, confirmamos nossa tese de que notícias verdadeiras são mais elaboradas utilizando mais palavras, adjetivos e um padrão de comprimento nos textos, assim como nos títulos. Sendo assim visto a relação com o diagrama de caixa de notícias confiáveis, observamos que textos possuem mais informações, contrariando as fake news que são compostas de títulos pequenos com objetivo de fornecer informações objetivas ao público alvo. Os números levantados no gráfico nos orientam que a mediana das notícias verdadeiras remete ao número do terceiro quartil relacionada as notícias falsas, mostrando a diferença de escalonamento entre o conteúdo dos textos.

Diagrama de Caixa da Quantidade de Palavras dos Textos

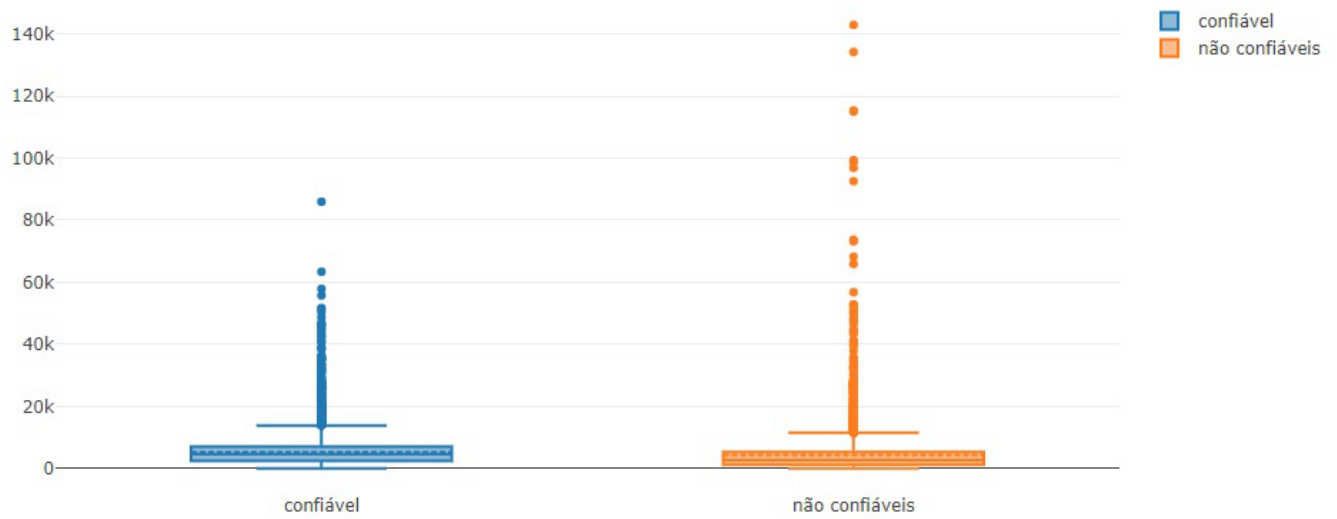
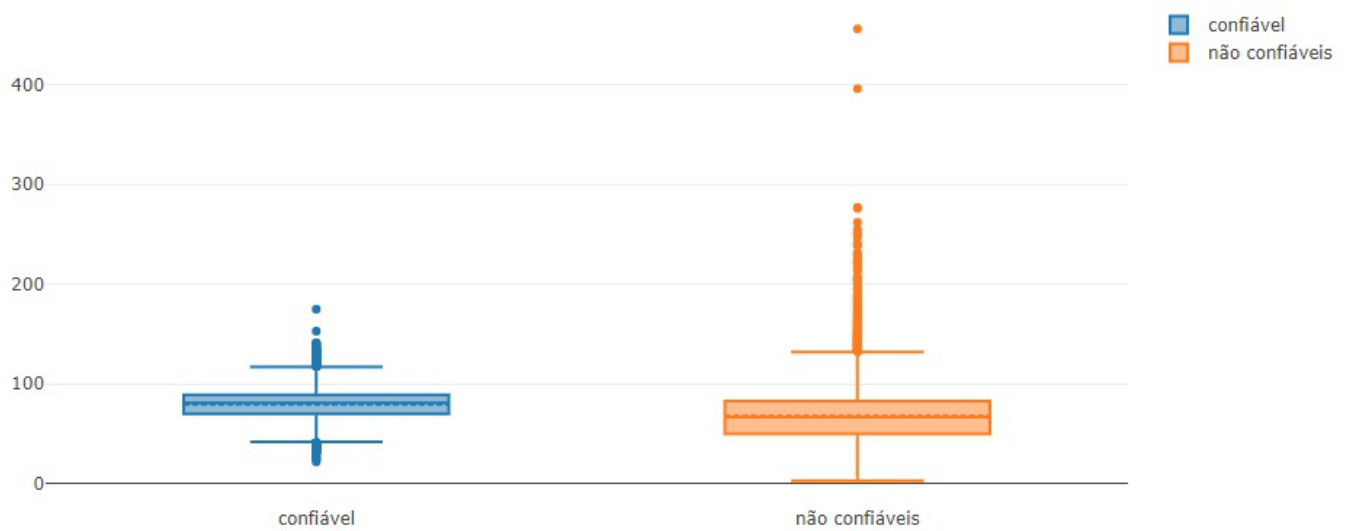


Diagrama de Caixa da Quantidade de Palavras dos Títulos

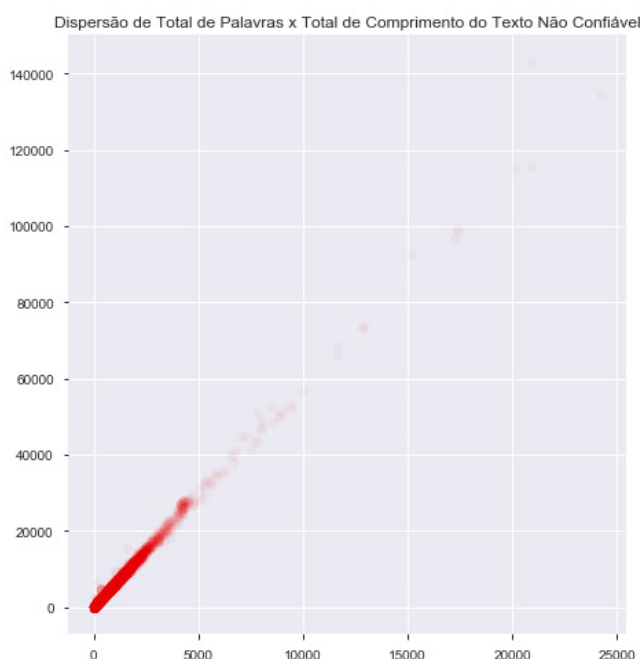
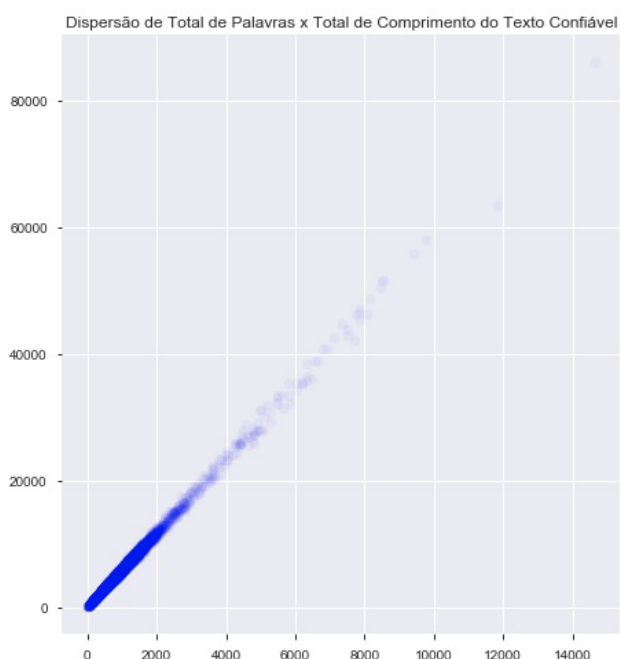


4.4 Dispersão

Gráfico de Dispersão é utilizado para pontuar dados em um eixo vertical e horizontal com a intenção de exibir quanto uma variável é afetada por outra. Existem vários tipos de associações entre parâmetros que podem ser demonstradas pelo gráfico de dispersão. A relação pode ser positiva ou negativa (quando um cresce o outro decresce), fraca ou forte, linear ou não linear.

A seguir, os gráficos abaixo representam a dispersão de total de palavras x total de comprimento do texto confiável e não confiável. A relação pode ser devida a outro parâmetro que esteja associado a cada um dos parâmetros estudados.

A associação demonstrada influencia na confiança do texto, quando o texto apresenta uma estrutura maior no conteúdo a chance de ser uma notícia verdadeira é maior, desta maneira podemos decifrar que o método das fake news é almejar informar a notícia de forma curta e rápida.



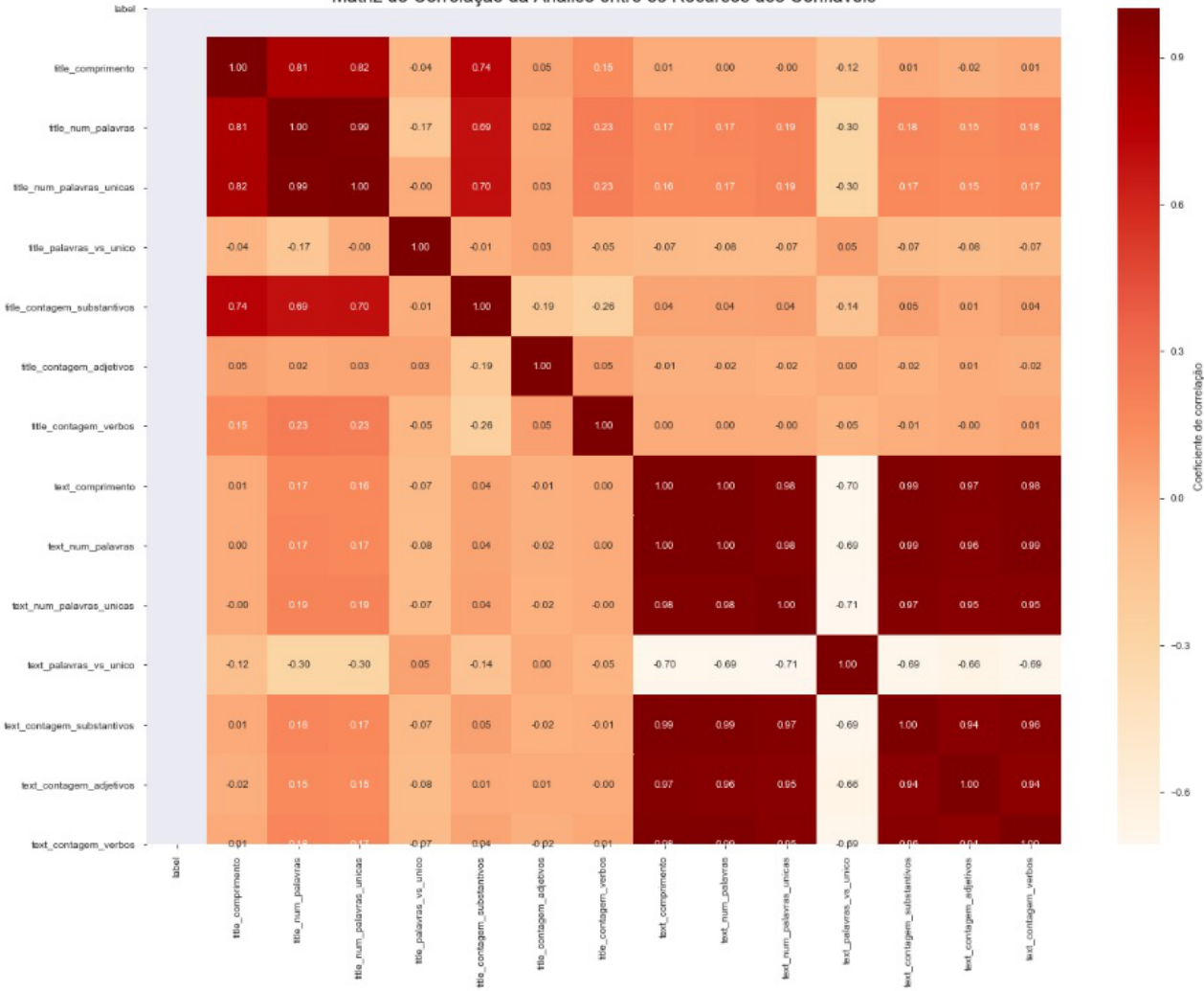
4.5 Correlação

A matriz de correlação mostra os valores de correlação de Pearson, que medem o grau de relação linear entre cada par de itens ou variáveis. Os valores de correlação podem cair entre -1 e +1. Uma vez feita a segmentação das variáveis, pudemos efetuar uma verificação de correlação mais concisa entre as variáveis de cada segmento, assim obtendo uma compreensão dos dados positivamente e negativamente relacionadas.

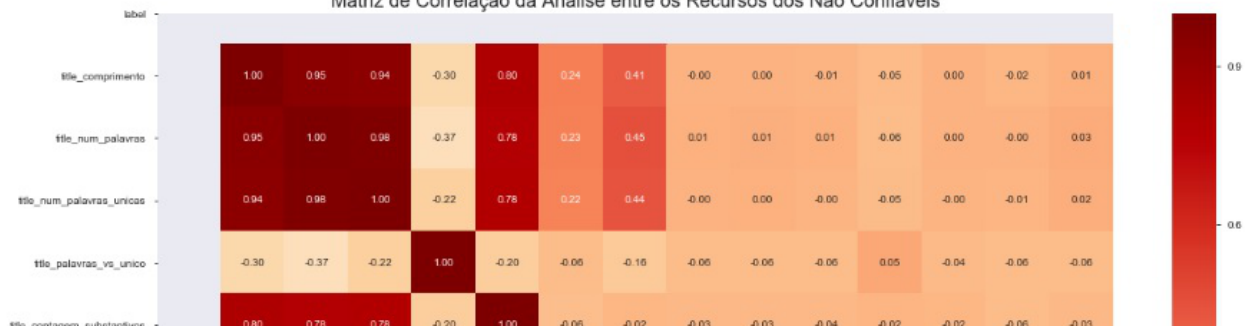
Com as variáveis extraídas por meio da análise por palavra, se tornou possível criar relacionamentos e identificar quais desses relacionamentos se correlacionam entre si.

Não identificamos um fato relevante no levantamento das correlações além havíamos constatado nos gráficos anteriores. Existe uma forte correlação entre texto e as classes gramaticais verbos e adjetivos, caso que não ocorre nos títulos, que correlaciona melhor com substantivos. A quantidade de palavras também tem forte correlação tanto nos títulos quanto nos textos.

Matriz de Correlação da Análise entre os Recursos dos Confiáveis



label



5 IMPLEMENTAÇÃO

Na etapa de aprendizado de máquina, decidimos utilizar um modelo em Deep Learning para treinar uma rede LSTM (Long Short Term Memory), uma variação de rede neural recorrente (RNN) utilizado em processamento de linguagem natural. Utilizaremos a biblioteca Keras TensorFlow para a criação desse modelo.

Inicialmente separamos os dados em treino e teste, o conjunto de dados que contém 15.848 registros, foi dividido em 80% de treinamento e 20% de teste, sendo assim, 12.678 registros para treinamento e 3.170 registros para testes.

Na tokenização das palavras, definiu-se um dicionário com no máximo 10.000 palavras. As sequências de entrada foram truncadas para ter o mesmo comprimento de 10.000, isso é necessário para executar o cálculo em Keras.

O modelo terá duas camadas LSTM, sendo uma com 128 e a outra com 64 unidades de memória (**LSTM(128)**), (**LSTM(64)**) e adicionado uma camada incorporada que utiliza 32 vetores por palavras (**embedding_vector_length = 32**). Para o Keras realizar o treinamento com duas camadas LSTM é necessário utilizar o parâmetro (**return_sequences = True**) responsável por permitir a conectividade entre camadas LSTM.

A escolha de 128 e 64, se refere sobre o número de células utilizadas, pois há um equilíbrio nos ajustes, resolvendo um problema conhecido como vanishing gradient, o qual ocorre quando os pesos computados nas partes iniciais da sequência perdem influência ao longo das iterações, sendo sensíveis às novas entradas, porém pode sofrer com sobre ajuste (overfitting).

Como se trata de redes neurais recorrentes, geralmente têm o problema de sobre ajuste (overfitting), para diminuir esse problema foi utilizada técnica de eliminação (dropout) em que os neurônios selecionados aleatoriamente são ignorados durante o treinamento. A ativação do neurônio é removida na passagem para frente e quaisquer atualizações de peso não são aplicadas ao neurônio na passagem para trás. Adicionamos três camadas de eliminação com 30% (**dropout = 0.3**).

Em nosso estudo enfrentamos um problema de classificação, queremos descobrir se determinado texto é Confiável ou Não Confiável. Para realizar as previsões 0 ou 1 para as classes (0 - Não confiável e 1 - Confiável) adicionou-se uma camada densa com 64 neurônios intermediários (**Dense(64, activation='relu')**) que nos ajuda a lidar com as saídas não-lineares, e um único neurônio de saída com função de ativação sigmoide (**Dense(1, activation='sigmoid')**) que classificará binariamente a saída, referenciada(1,).

Para analisar os resultados utilizamos as métricas logloss e acurácia e para compilação o algoritmo de otimização do ADAM (**loss='binary_crossentropy', optimizer='adam', metrics=['accuracy']**).

O treinamento foi realizado em 3 épocas com lote de 64 classificações para espaçar as atualizações de pesos (**X_train_seq, y_train, nb_epoch=3, batch_size=64, validation_data=(X_test_seq,y_test)**)

5.1 Modelo em Keras TensorFlow

Imports

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Dropout
from keras.layers.embeddings import Embedding
from keras.preprocessing import sequence
from keras.preprocessing.text import Tokenizer
```

Recebendo Conjunto de Dados

```
path_train = 'train.csv'
train_data = pd.read_csv(path_train, encoding='utf-8')
```

Criando novos recursos

```
train_data['title_author_text'] = train_data['title'] + ' ' + train_data['author'] + ' ' + train_data['text']
train_data['len_title_author_text'] = [len(str(x)) for x in train_data['title_author_text']]
```

Iniciando Modelo em Keras TensorFlow

```
np.random.seed(7)
```

```
train_features = train_data['title_author_text']
train_targets = train_data['label']
```

```
X_train, X_test, y_train, y_test = train_test_split(train_features, train_targets, test_size=0.2, random_state=42)
```

```
print('Dados de Treino - Feature: {}'.format(len(X_train)))
print('Dados de Treino - Label: {}'.format(len(y_train)))
print(' ')
print('Dados de Teste - Feature: {}'.format(len(X_test)))
print('Dados de Teste - Label: {}'.format(len(y_test)))
```

6 RESULTADO FINAL

Após construir o modelo de treinamento, foi possível demonstrar os resultados conforme nosso Kernel, onde obtivemos resultados de 80% de acerto em nossos testes, consideramos um resultado satisfatório para o início da pesquisa.

Durante a criação do modelo realizamos diversas modelagens, inclusive, testamos outras formas de aprendizado de máquina, como : SVC, KNN, Árvore de Decisão, RandomForest e Gradient Boosting Classifier, porém, de acordo com as observações feitas, a rede neural LSTM apresentou melhor resultado para a nossa aplicação. Não descartamos que ao usar uma gama maior de notícias e atributos distintos de análise sintática e semântica a necessidade de modificações neste escopo. Porém, em alguns métodos como de soma de sufixo, onde se soma sempre o vetor da próxima palavra observamos problemas em relação a tempo de execução. Infelizmente por falta de tempo hábil não foi possível representar esta funcionalidade neste trabalho.

Mas como visto foi possível provar que através de um mapa auto organizável de uma Rede Neural Recorrente podemos separar textos através de seus padrões de entrada e gerar classificações semelhantes as apresentadas no córtex cerebral humano.

Com base nos critérios e resultados do presente estudo, onde o objetivo era apresentar uma maneira de classificar notícias falsas desenvolvida por meio de uma Redes Neural, concluímos que o objetivo foi alcançado de maneira satisfatória. Através dos conceitos aplicados e nos agrupamentos das informações, foi possível verificar que ao apresentar um padrão de limpeza nos dados, associação entre termos, padrões verbais, o modelo conseguiu um resultado bem-sucedido na classificação, de forma que será possível classificar uma informação sendo como notícia falsa ou não com chance de acerto em 75.00%.

Para que isso fosse possível as ferramentas utilizadas foram a linguagem Python e biblioteca Keras TensorFlow utilizando a função de redes neurais LSTM, por fim todos os processos foram executados com sucesso.

O que reforça a afirmação feita no início do trabalho, quando classificou que redes neurais recorrentes permitem um reconhecimento padrões de texto, sempre com uma alta taxa de assertividade.

REFERÊNCIAS

- [1] DAVIDSON, Thomas et al. Hate Speech and Offensive Language: 24k tweets labeled as hate speech, offensive language, or neither. 2017. Disponível em: <<https://data.world/thomasrdavidson/hate-speech-and-offensive-language>>. Acesso em: 14 set. 2018.
- [2] SANTANA, Felipe. Fake news, notícias em texto e oportunidades. 2018. Disponível em: <<http://minerandodados.com.br/index.php/2018/10/09/fake-news-noticias-em-texto-e-oportunidades/>>. Acesso em: 12 out. 2018.
- [3] PEREZ-ROSAS, Verônica et al. Automatic Detection of Fake News. 2016. 10 f. Department of Psychology (Computer Science and Engineering,) - University of Michigan, [S.I.], 2016.
- [4] W. PENNEBAKER, James. Lying Words: Predicting Deception From Linguistic Styles. In: L. NEWMAN, Mathew et al. Lying Words: Predicting Deception From Linguistic Styles. 2003. ed. [S.I.]: PSPB, 2009. p. 665-675. v. 1.
- [5] PEREZ-ROSAS, Verônica; MIHALCEA, Rada. Experiments in Open Domain Deception Detection. 2008. 6 p. Computer Science (Computer Science and Engineering) - University of Michigan, [S.I.], 2008.
- [6] DA ROCHA, Bernardo Abbad et al. O avanço das fake news e sua retratação na mídia de referência. 2018. 15 p. Sociedade Brasileira de Estudos Interdisciplinares da Comunicação (Congresso de Ciências da Comunicação na Região Sul) - Universidade Federal de Santa Maria, Santa Maria, RS, Cascavel-PR, 2018.
- [7] RUBIN, Victoria L. et al. Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. 2015. 11 p. Language and Information Technology Research Lab (Faculty of Information and Media Studies) - University of Western Ontario, London, Ontario, CANADA, 2015.
- [8] SHAFQAT, Wafa et al. The Language of Deceivers: Linguistic Features of Crowdfunding Scams. 2015. 2 p. Sangmyung University (Sangmyung University)- Sangmyung University, Cheonan, South Korea, 2015.
- [9] CONROY, Niall J.; RUBIN, Victoria L.; CHEN, Yimin. Automatic Deception Detection: Methods for Finding Fake News. 2014. 4 f. Language and Information Technology Research Lab (LIT.RL) (Faculty of Information and Media Studies) - University of Western Ontario, London, Ontario, CANADA, 2014.
- [10] DAVIDSON, Thomas et al. Automated Hate Speech Detection and the Problem of Offensive Language. 2013. 4 f. Department of Applied Mathematics (Qatar Computing Research Institute) - Cornell University, 3Department of Information Science, Ithaca, NY, USA, 2013.
- [11] MIN H. Kao Department of Electrical Engineering & Computer Science. [S. I.], 2019. Disponível em: <https://www.eecs.utk.edu/people/student-organizations/>. Acesso em: 9 jul. 2019.
- [12] BROWNLEE , Jason. A Gentle Introduction to the Rectified Linear Unit (ReLU). [S. I.], 9 jan. 2019. Disponível em: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/>. Acesso em: 9 jul. 2019.
- [13] DEPARTMENT of Electrical Engineering and Computer Science - UTK. [S. I.], 2019. Disponível em: <https://www.facebook.com/EECS.UTK/>. Acesso em: 9 jul. 2019.
- [14] BROWNLEE, Jason. Dropout Regularization in Deep Learning Models With Keras. [S. I.], 20 jun. 2016. Disponível em: <https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/>. Acesso em: 9 jul. 2019.
- [15] BIDIRECTIONAL Long Short-Term Memory (BI-LSTM) with Attention Mechanism. [S. I.], 30 jun. 2019. Disponível em: [http://primo.ai/index.php?title=Bidirectional_Long_Short-Term_Memory_\(BI-LSTM\)_with_Attention_Mechanism](http://primo.ai/index.php?title=Bidirectional_Long_Short-Term_Memory_(BI-LSTM)_with_Attention_Mechanism). Acesso em: 9 jul. 2019.

- [16] BROWNLEE, Jason. Sequence Classification with LSTM Recurrent Neural Networks in Python with Keras. [S. l.], 26 jul. 2016. Disponível em: <https://machinelearningmastery.com/sequence-classification-lstm-recurrent-neural-networks-python-keras/>. Acesso em: 9 jul. 2019.
- [17] BIDIRECTIONAL recurrent neural networks. [S. l.], 7 jul. 2019. Disponível em: https://en.wikipedia.org/wiki/Bidirectional_recurrent_neural_networks. Acesso em: 9 jul. 2019.
- [18] A BEGINNER'S Guide to LSTMs and Recurrent Neural Networks. [S. l.], 2019. Disponível em: <https://skymind.ai/wiki/lstm>. Acesso em: 9 jul. 2019.
- [19] OLAH, Christopher. Understanding LSTM Networks. [S. l.], 27 ago. 2015. Disponível em: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Acesso em: 9 jul. 2019.
- [20] HE, Lang He et al. Multimodal Affective Dimension Prediction Using Deep Bidirectional Long Short-Term Memory Recurrent Neural Networks. Proceeding AVEC '15 Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge Pages 73-80, 26 out. 2015. Disponível em: <https://dl.acm.org/citation.cfm?id=2811641>. Acesso em: 9 jul. 2019.
- [21] GUAN, Zhonghui et al. Integration of Long-Term-Memory-Related SynapticPlasticity Involves Bidirectional Regulationof Gene Expression and Chromatin Structure. [S. l.], 15 nov. 2002. Disponível em: <https://reader.elsevier.com/reader/sd/pii/S0092867402010747?token=73DB87FE6A1527FD3EF86124A1AC74C4D12672A129141DEEE0E768066DC0E8118D7DE4A039BB31A75FE45A3EE126C6C3>. Acesso em: 9 jul. 2019.
- [22] LIWICKI, Marcus et al. A Novel Approach to On-Line Handwriting Recognition Based on Bidirectional Long Short-Term Memory Networks. [S. l.], 2019. Disponível em: <https://mediatum.ub.tum.de/doc/1289961/file.pdf>. Acesso em: 9 jul. 2019.
- [23] ZHOU, Peng et al. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. [S. l.], 2019. Disponível em: <https://www.aclweb.org/anthology/P16-2034>. Acesso em: 9 jul. 2019.
- [24] Frei , Lukas. Speed Up Your Exploratory Data Analysis With Pandas-Profiling. [S. l.], 25 apr. 2019. Disponível em: <https://towardsdatascience.com/speed-up-your-exploratory-data-analysis-with-pandas-profiling-88b33dc53625>. Acesso em: 9 jul. 2019.
- [25] Lucendo, Guillermo Altares. A longa história das notícias falsas, 18 jun 2018. Disponível em: https://brasil.elpais.com/brasil/2018/06/08/cultura/1528467298_389944.html. Acesso em: 01 out. 2019.
- [26] Réflexions d'un historien sur les fausses nouvelles de la guerre. Disponível em: https://fr.wikisource.org/wiki/R%C3%A9flexions_d%E2%80%99un_historien_sur_les_fausses_nouvelles_de_la_guerre. Acesso em: 01 out. 2019.
- [27] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts and Jonathan L. Zittrain. The science of fake news. Disponível em: https://scholar.harvard.edu/files/mbaum/files/science_of_fake_news.pdf. Acesso em: 01 out. 2019.
-