



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO  
DISCIPLINA APRENDIZADO PROFUNDO

TESTES COM MODELOS DE APRENDIZADO PROFUNDO  
(TRABALHO FINAL)

ALUNA:  
VIVIAN RIQUE GIL FERRARO

RIO DE JANEIRO– BRASIL  
JULHO DE 2023

## SUMÁRIO

1.	Introdução .....	2
2.	Desenvolvimento.....	2
2.1.	Bancos de dados utilizados.....	2
2.2.	Modelos.....	3
3.	Resultados .....	4
3.1.	Modelo LSTM com conjunto de dados ISOT Fake News Dataset .....	4
3.2.	Modelo LSTM com conjunto de dados WELFAKE.....	8
3.3.	Modelo Bert com conjunto de dados ISOT Fake News Dataset .....	12
3.4.	Modelo Bert com conjunto de dados WELFAKE .....	15
4.	Conclusões .....	19
5.	Referências .....	20

## 1. Introdução

As notícias falsas são um problema crescente nas sociedades e têm o potencial de influenciar a opinião pública, espalhar desinformação e até mesmo afetar os resultados das eleições. É importante desenvolver ferramentas eficazes para identificar e combater as notícias falsas. Alguns modelos de aprendizado profundo, como os baseados em Redes Neurais Recorrentes (RNNs), ou em redes Transformers, são muito eficientes no processamento de linguagem natural, e podem ser usados para identificar notícias falsas com alta precisão.

Neste trabalho testei alguns modelos de aprendizado profundo para classificar um conjunto de notícias pertencentes a um banco de dados pré-rotulado. Os modelos foram treinados no conjunto de dados e, em seguida, utilizados para classificar as notícias como verdadeiras ou falsas. Ao avaliar o desempenho dos modelos, foi possível identificar quais abordagens foram mais eficazes na identificação de notícias falsas.

## 2. Desenvolvimento

### 2.1. Bancos de dados utilizados

Para o trabalho utilizei dois bancos de dados:

#### 1. ISOT Fake News Dataset (Disponível em:

<https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/>)

Nesse conjunto, os dados estão dispostos em dois arquivos do tipo .csv.

O primeiro arquivo contém um conjunto com 21.417 notícias verdadeiras, e o segundo 23.481 notícias falsas, após tratamento, onde foram retiradas algumas entradas não rotuladas, os arquivos foram concatenados e convertidos em um arquivo do Excel, contendo 44.895 entradas.

#### 2. WELFake Dataset (Disponível em:

<https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification>

O dataset está contido em um único arquivo com 72.135 entradas, sendo 35.028 notícias verdadeiras e 37.106 notícias falsas. O arquivo .csv foi convertido em arquivo do Excel.

## 2.2. Modelos

### BERT

Ou ‘Bidirectional Encoder Representations from Transformers’ é um modelo de linguagem pré-treinado que foi desenvolvido pelo Google AI em 2018. É um modelo de pré-treinamento de processamento de linguagem natural, ele foi treinado em um grande conjunto de dados de texto, como a Wikipédia, para aprender a representar a linguagem humana. Isso permite que o modelo aprenda as relações entre as palavras e o contexto em que elas são usadas.

A principal inovação técnica do BERT é a aplicação do treinamento bidirecional do Transformer (um modelo de atenção) à modelagem de linguagem. Isso significa que, em vez de olhar para uma sequência de texto em uma única direção (da esquerda para a direita ou da direita para a esquerda), o BERT é capaz de analisar o contexto de uma palavra olhando para as palavras à sua esquerda e à sua direita ao mesmo tempo.

### LTSM

Ou ‘Long Short-Term Memory’ é um tipo de rede neural recorrente (RNN) que é usada para sequências de dados. LSTMs são capazes de aprender padrões em sequências de dados de longa duração porque usam uma estrutura de memória interna que permite que eles guardem informações sobre as entradas anteriores por um período de tempo. Essa estrutura de memória é chamada de "unidade de memória de longo prazo.

A LSTM é uma rede neural recorrente que possui uma estrutura em cadeia. A estrutura em cadeia é composta por quatro redes neurais, que são chamadas de células. Cada célula é responsável por armazenar informações sobre uma entrada específica. As células são então conectadas entre si, o que permite que a LSTM aprenda padrões em sequências de dados de longa duração.

### 3. Resultados

Foram feitos alguns testes utilizando os conjuntos de dados e os modelos citados, ao longo dessas experimentações, alguns hiper parâmetros foram reajustados, buscando observar o impacto nos resultados.

Alguns desses experimentos estão descritos abaixo.

#### 3.1. Modelo LSTM com conjunto de dados ISOT Fake News Dataset

##### 3.1.1 Treino 1

Código disponível em:

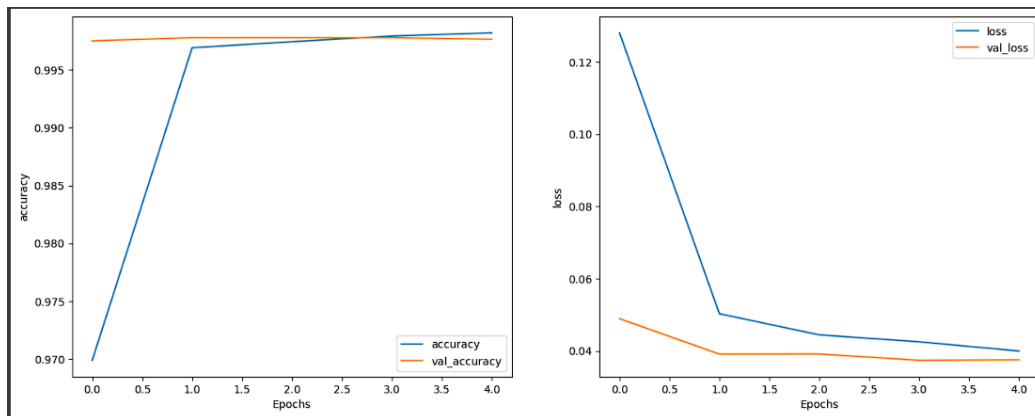
<https://github.com/viferraro/RedesFakeNews/blob/main/ISOTLSTM.ipynb>

- Características principais da rede:
  - Rede sequencial.
  - A primeira camada é uma camada de Embedding que converte os índices das palavras em vetores densos de tamanho fixo. A camada tem como parâmetros: o tamanho do vocabulário (1000), a dimensão do embedding (25) e o comprimento máximo das sequências (100).
  - A segunda camada é uma camada de Dropout com taxa de dropout de 0.5, usada para prevenir overfitting durante o treinamento.
  - A terceira camada é uma camada Bidirectional com uma camada LSTM com 16 unidades dentro.
  - A quarta camada é outra camada de Dropout com taxa de dropout de 0.5.
  - A quinta e última camada é uma camada Densa com 1 unidade, ativação sigmoide e regularização L2 nos pesos (0.01) e bias (0.01) da camada. Essa camada produz a saída final da rede, representando a probabilidade da classe positiva.
  - O modelo é compilado com perda `binary_crossentropy`, otimizador `adam` e métrica `acurácia`.
- Total de Parâmetros: 30.409
- Batch-size: 16
- Epochs: 10

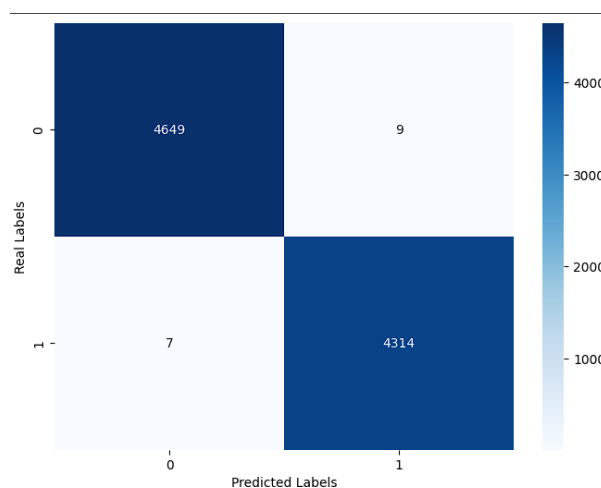
- Utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

O treino foi interrompido após 5 épocas, pela configuração do método EarlyStopping, quando os valores da acurácia de validação não variassem significativamente após 3 épocas consecutivas.

O modelo parece ter aprendido bem e bem rápido, pois podemos notar que os valores de treino e validação seguem a mesma tendência.



A matriz de confusão gerada também sugere um ótimo desempenho do modelo, com 4.649 verdadeiros negativos, 7 falsos negativos, 4.314 verdadeiros positivos e apenas 9 falsos positivos.



Resumo do desempenho:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4658
1	1.00	1.00	1.00	4321
accuracy			1.00	8979
macro avg	1.00	1.00	1.00	8979
weighted avg	1.00	1.00	1.00	8979

### 3.1.2 Treino 2

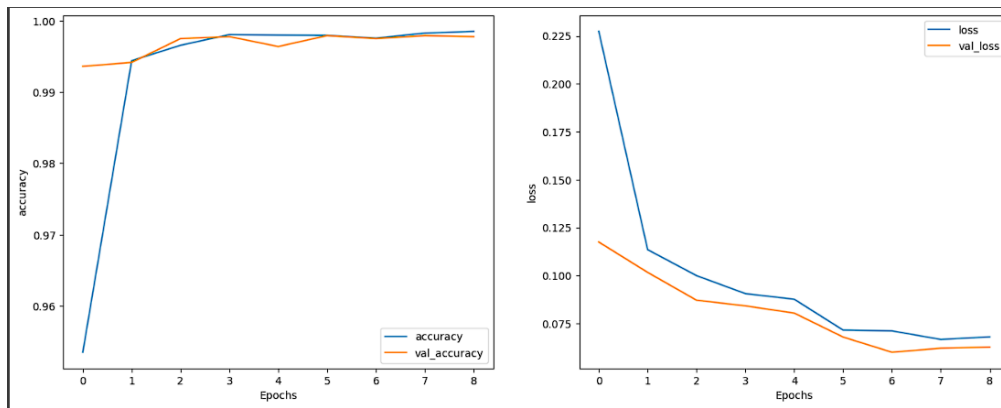
Código disponível em:

<https://github.com/viferraro/RedesFakeNews/blob/main/ISOTLSTM2.ipynb>

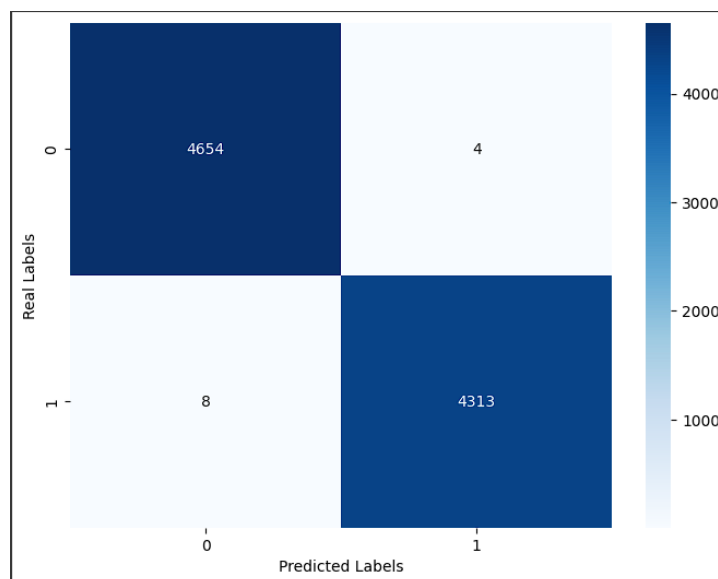
- Características principais da rede:
  - Rede sequencial.
  - A primeira camada é uma camada de Embedding , que tem como parâmetros: o tamanho do vocabulário (1000), a dimensão do embedding (50) e o comprimento máximo das sequências (100).
  - A segunda camada é uma camada de Dropout com taxa de dropout de 0.5, usada para prevenir overfitting durante o treinamento.
  - A terceira camada é uma camada Bidirectional com uma camada LSTM com 32 unidades dentro.
  - A quarta camada é outra camada de Dropout com taxa de dropout de 0.5.
  - A quinta e última camada é uma camada Dense com 1 unidade, ativação sigmoide e regularização L2 nos pesos (0.1) e bias (0.1) da camada. Essa camada produz a saída final da rede, representando a probabilidade da classe positiva.
  - O modelo é compilado com perda `binary_crossentropy`, otimizador `adam` e métrica `acurácia`.
- Total de Parâmetros: 71.313
- Batch-size: 32
- Epochs: 10
- Utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

O treino foi interrompido após 9 épocas, pela configuração do método `EarlyStopping`. Nesse segundo teste, eu alterei o número de neurônios da camada LSTM (32), a dimensão do vetor de entrada (50) e o tamanho do batch (32).

O modelo continua aparentemente aprendendo bem e rápido, os valores de treino e validação ainda seguem a mesma tendência.



A matriz de confusão gerada também sugere um ótimo desempenho do modelo, com 4.654 verdadeiros negativos, 8 falsos negativos, 4.313 verdadeiros positivos e apenas 4 falsos positivos.



Resumo do desempenho:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4658
1	1.00	1.00	1.00	4321
accuracy			1.00	8979
macro avg	1.00	1.00	1.00	8979
weighted avg	1.00	1.00	1.00	8979



### 3.2. Modelo LSTM com conjunto de dados WELFAKE

Resolvi testar os modelos com um conjunto de dados um pouco maior, para avaliar se o modelo continuará aprendendo rápido.

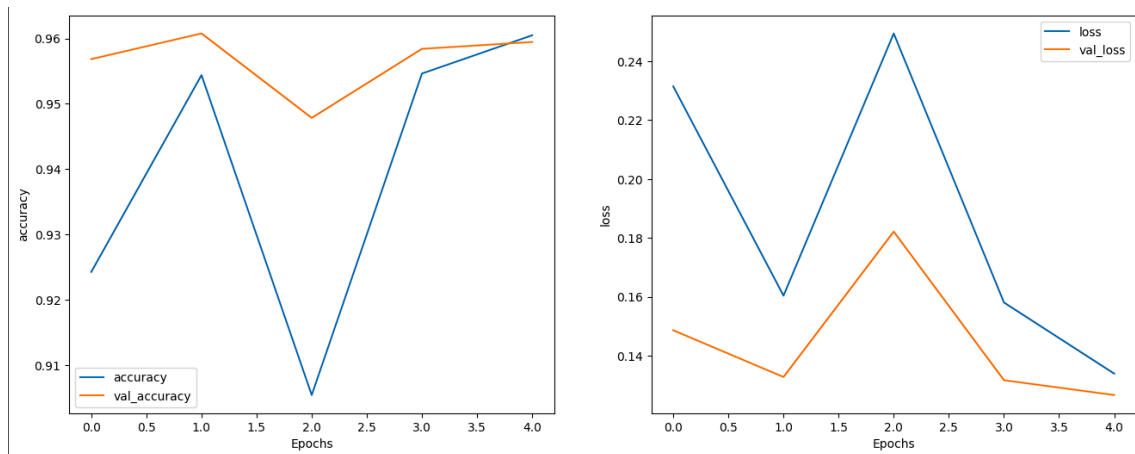
#### 3.2.1 Teste 1

Código disponível em:

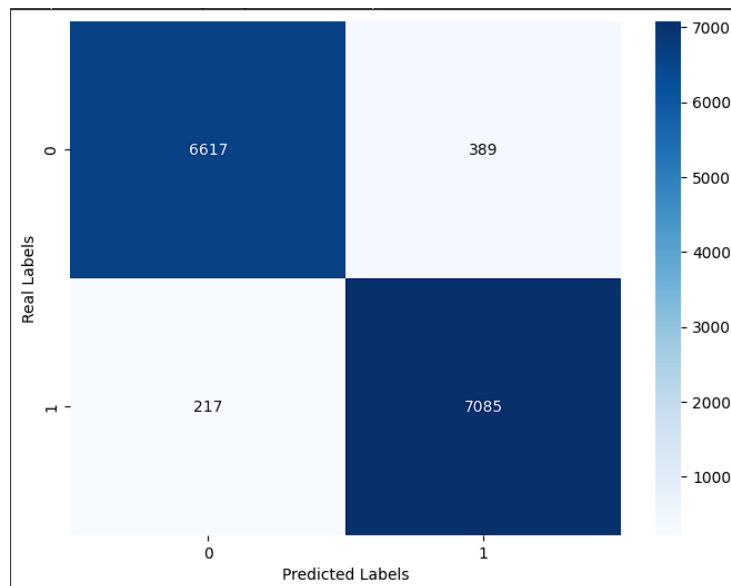
- <https://github.com/viferraro/RedesFakeNews/blob/main/WELFAKELSTM16.ipynb>
- Características principais da rede:
  - Rede sequencial.
  - A primeira camada é uma camada de Embedding , que tem como parâmetros: o tamanho do vocabulário (2000), a dimensão do embedding (25) e o comprimento máximo das sequências (100).
  - A segunda camada é uma camada de Dropout com taxa de dropout de 0.5, usada para prevenir overfitting durante o treinamento.
  - A terceira camada é uma camada Bidirectional com uma camada LSTM com 16 unidades dentro.
  - A quarta camada é outra camada de Dropout com taxa de dropout de 0.5.
  - A quinta e última camada é uma camada Dense com 1 unidade, ativação sigmoide e regularização L2 nos pesos (0.01) e bias (0.01) da camada. Essa camada produz a saída final da rede, representando a probabilidade da classe positiva.
  - O modelo é compilado com perda `binary_crossentropy`, otimizador `adam` e métrica `acurácia`.
- Total de Parâmetros: 55.409
- Batch-size: 16
- Epochs: 10
- Utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

O treino foi interrompido após 5 épocas, pela configuração do método `EarlyStopping`. O modelo parece aprender bem, mas notamos uma entre a acurácia no conjunto de treinamento e no conjunto de validação, que começa aumentar, passando de 0.0325 na primeira época para 0.0390 na quinta época. Isso pode ser um sinal de overfitting,

ou seja, o modelo pode estar começando a memorizar os dados de treinamento em vez de aprender a generalizar para novos dados.



A matriz de confusão sugere que o modelo tem um bom desempenho, com 6.617 verdadeiros negativos, 389 falsos negativos, 7.085 verdadeiros positivos e 217 falsos positivos.



Resumo do desempenho:

	precision	recall	f1-score	support
0	0.97	0.94	0.96	7006
1	0.95	0.97	0.96	7302
accuracy			0.96	14308
macro avg	0.96	0.96	0.96	14308
weighted avg	0.96	0.96	0.96	14308

### 3.2.2 Teste 2

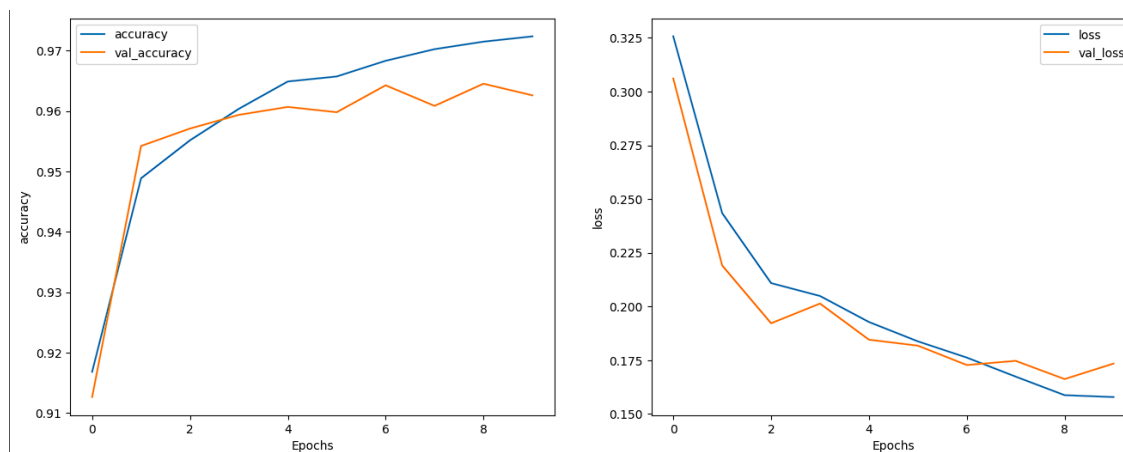
Código disponível em:

- [https://github.com/viferraro/RedesFakeNews/blob/main/WELFAKELSTM16\\_2.ipynb](https://github.com/viferraro/RedesFakeNews/blob/main/WELFAKELSTM16_2.ipynb)
- Características principais da rede:
  - Rede sequencial.
  - A primeira camada é uma camada de Embedding , que tem como parâmetros: o tamanho do vocabulário (2000), a dimensão do embedding (50) e o comprimento máximo das sequências (100).
  - A segunda camada é uma camada de Dropout com taxa de dropout de 0.5, usada para prevenir overfitting durante o treinamento.
  - A terceira camada é uma camada Bidirectional com uma camada LSTM com 16 unidades dentro.
  - A quarta camada é outra camada de Dropout com taxa de dropout de 0.5.
  - A quinta e última camada é uma camada Dense com 1 unidade, ativação sigmoide e regularização L2 nos pesos (0.1) e bias (0.1) da camada. Essa camada produz a saída final da rede, representando a probabilidade da classe positiva.
  - O modelo é compilado com perda `binary_crossentropy`, otimizador `adam` e métrica `acurácia`.
- Total de Parâmetros: 108.609
- Batch-size: 16
- Epochs: 10
- Utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

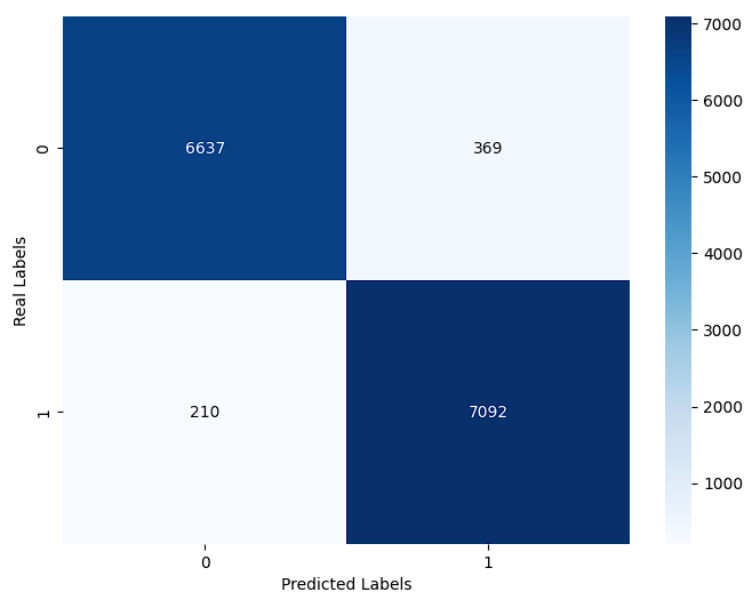
O treino não foi interrompido.

Nesse segundo teste, eu alterei o tamanho do vocabulário (2000), a dimensão do vetor de entrada (50) e a regularização L2 (0.1).

O modelo teve um desempenho muito bom, com valores de treino e validação seguindo a mesma tendência.



A matriz de confusão gerada mostra um bom desempenho, com 6.637 verdadeiros negativos, 210 falsos negativos, 7.092 verdadeiros positivos e 369 falsos positivos.



Resumo do desempenho:

	precision	recall	f1-score	support
0	0.97	0.95	0.96	7006
1	0.95	0.97	0.96	7302
accuracy			0.96	14308
macro avg	0.96	0.96	0.96	14308
weighted avg	0.96	0.96	0.96	14308

### 3.3. Modelo Bert com conjunto de dados ISOT Fake News Dataset

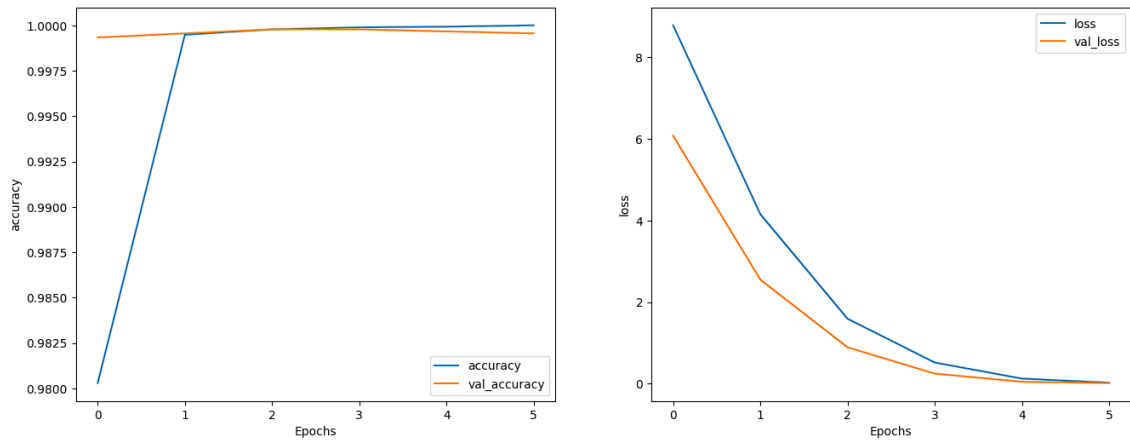
#### 3.3.1 Teste 1

Código disponível em:

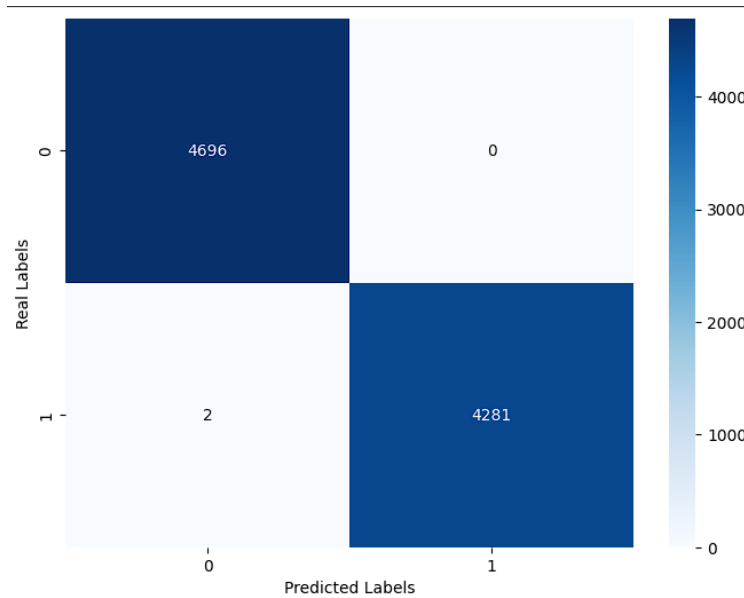
<https://github.com/viferraro/RedesFakeNews/blob/main/ISOTBert.ipynb>

- Características principais da rede:
  - Rede do tipo Model.
  - A rede tem duas entradas: `input_ids` e `input_mask`, ambas com formato (MAX\_LEN). Essas entradas representam os IDs dos tokens e as máscaras de atenção, respectivamente, que são usados como entrada para o modelo BERT.
  - A saída do pooler é usada como entrada para a próxima camada da rede. Essa saída é um vetor de tamanho fixo que representa a sequência de entrada e é passada por uma camada de Dropout com taxa de dropout de 0.5.
  - Em seguida, a saída da camada de Dropout é passada por uma camada Dense com 64 unidades, ativação ReLU e regularização L2 (0.1) nos pesos e L1 (0.1) nos bias da camada.
  - A saída da camada Dense é passada por outra camada de Dropout com taxa de dropout de 0.5.
  - A última camada é uma camada Dense com 1 unidade e ativação sigmoide, usada para produzir a saída final da rede, representando a probabilidade da classe positiva.
  - O modelo é compilado com otimizador Adam, taxa de aprendizado 1e-05, perda `binary_crossentropy` e métrica acurácia.
- Total de Parâmetros: 109.531.521
- Batch-size: 16
- Epochs: 10
- Utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

O treino foi interrompido após 6 épocas, pela configuração do método EarlyStopping. O modelo teve um aparente bom desempenho, com valores de treino e validação seguindo a mesma tendência.



A matriz de confusão sugere um desempenho quase perfeito, com 4.696 verdadeiros negativos, 2 falsos negativos, 4.281 verdadeiros positivos e nenhum falso positivo.



Resumo do desempenho:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4696
1	1.00	1.00	1.00	4283
accuracy			1.00	8979
macro avg	1.00	1.00	1.00	8979
weighted avg	1.00	1.00	1.00	8979

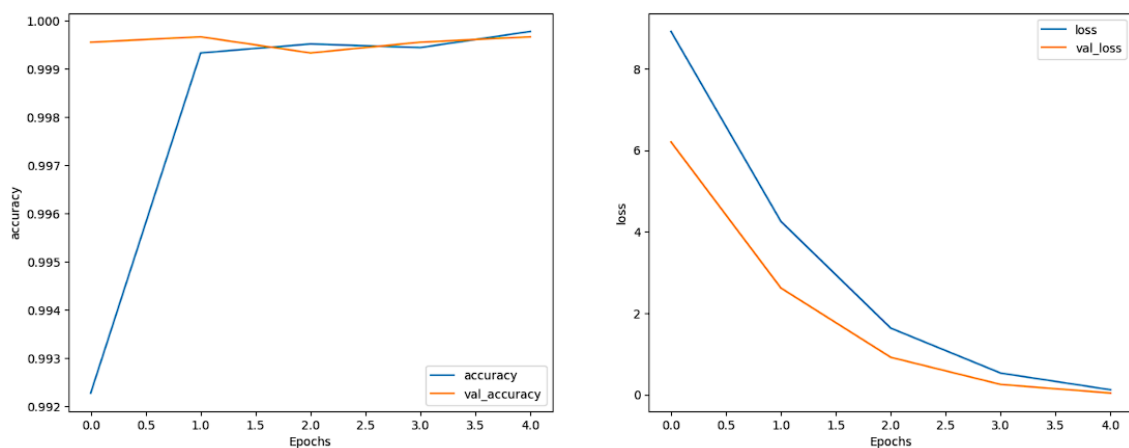
### 3.3.2 Teste 2

Código disponível em:

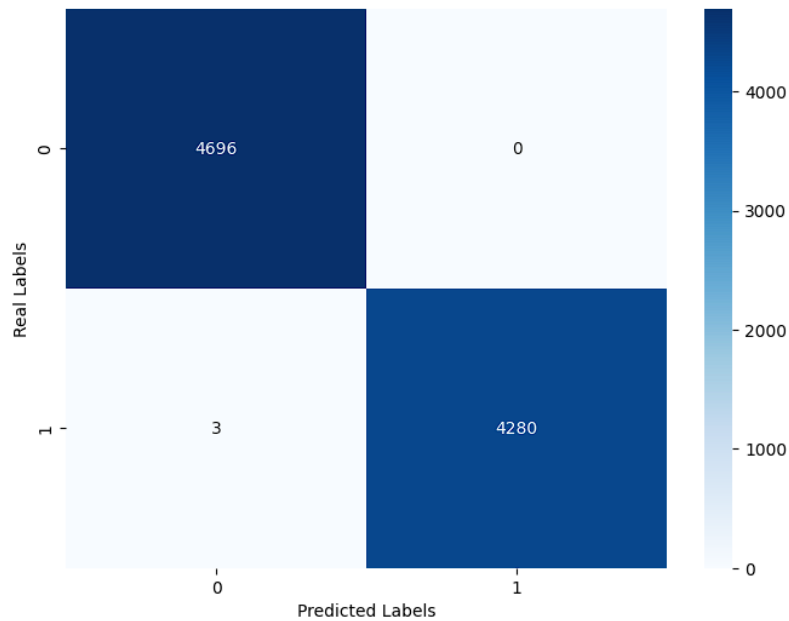
<https://github.com/viferraro/RedesFakeNews/blob/main/ISOTBertBatchN.ipynb>

- Características principais da rede:
  - Rede do tipo Model, sem muitas alterações em relação à anterior.
  - Foi introduzida uma camada de BatchNormalization após a camada densa..
- Total de Parâmetros: 109.531.777 (109.531.649 treináveis)
- Batch-size: 16
- Epochs: 10
- Utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

O treino foi interrompido após 6 épocas, pela configuração do método EarlyStopping. O modelo teve um aparente bom desempenho, com valores de treino e validação próximos.



A matriz de confusão também sugere um desempenho excelente, com 4.696 verdadeiros negativos, 3 falsos negativos, 4.280 verdadeiros positivos e nenhum falso positivo.



Resumo do desempenho:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4696
1	1.00	1.00	1.00	4283
accuracy			1.00	8979
macro avg	1.00	1.00	1.00	8979
weighted avg	1.00	1.00	1.00	8979

### 3.4. Modelo Bert com conjunto de dados WELFAKE

#### 3.4.1 Teste1

Código disponível em:

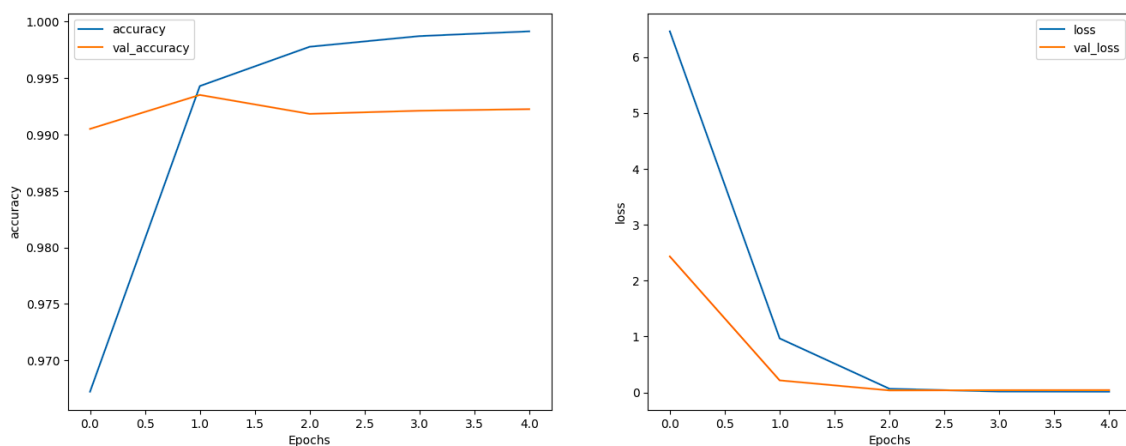
<https://github.com/viferraro/RedesFakeNews/blob/main/WelFakeBert.ipynb>

- Características principais da rede:
  - Rede do tipo Model.
  - A rede tem duas entradas: `input_ids` e `input_mask`, ambas com formato (MAX\_LEN). Essas entradas representam os IDs dos tokens e as máscaras de atenção, respectivamente, que são usados como entrada para o modelo BERT.
  - A saída do pooler é usada como entrada para a próxima camada da rede. Essa saída é um vetor de tamanho fixo que representa a sequência de entrada e é passada por uma camada de Dropout com taxa de dropout de 0.5.

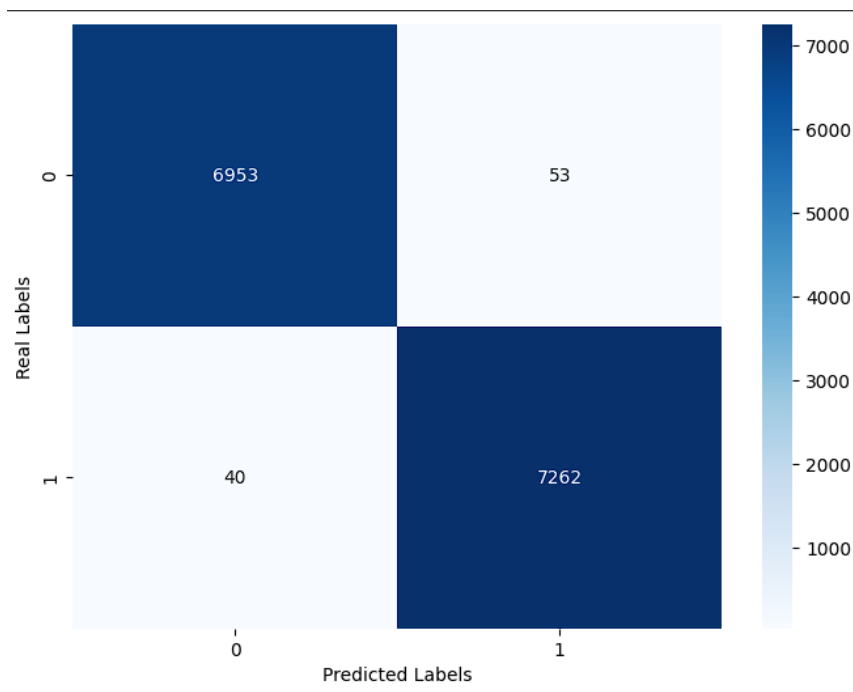


- Em seguida, a saída da camada de Dropout é passada por uma camada Dense com 64 unidades, ativação ReLU e regularização L2 (0.1) nos pesos e L1 (0.1) nos bias da camada.
  - A saída da camada Dense é passada por outra camada de Dropout com taxa de dropout de 0.5.
  - A última camada é uma camada Dense com 1 unidade e ativação sigmoide, usada para produzir a saída final da rede, representando a probabilidade da classe positiva.
  - O modelo é compilado com otimizador Adam, taxa de aprendizado  $1e-05$ , perda `binary_crossentropy` e métrica acurácia.
- Total de Parâmetros: 109.531.521
  - Batch-size: 16
  - Epochs: 10
  - Não utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

O treino foi interrompido após 5 épocas, pela configuração do método EarlyStopping. O modelo teve um aparente bom desempenho, com pouca diferença entre os valores dos conjuntos de treino e validação.



A matriz de confusão sugere um desempenho muito bom, com 6.953 verdadeiros negativos, 40 falsos negativos, 7.262 verdadeiros positivos e 53 falsos positivos.



Resumo do desempenho:

	precision	recall	f1-score	support
0	0.99	0.99	0.99	7006
1	0.99	0.99	0.99	7302
accuracy			0.99	14308
macro avg	0.99	0.99	0.99	14308
weighted avg	0.99	0.99	0.99	14308

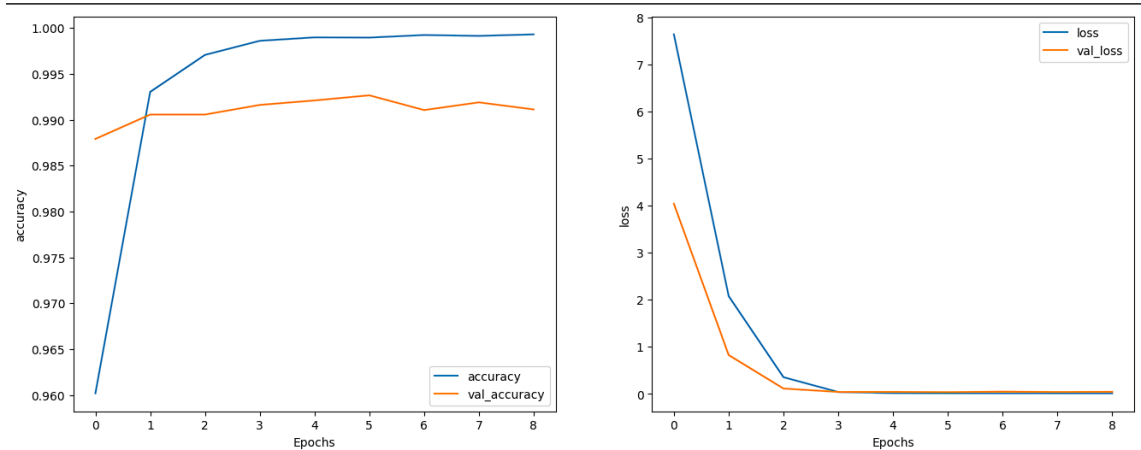
### 3.4.2 Teste 2

Código disponível em:

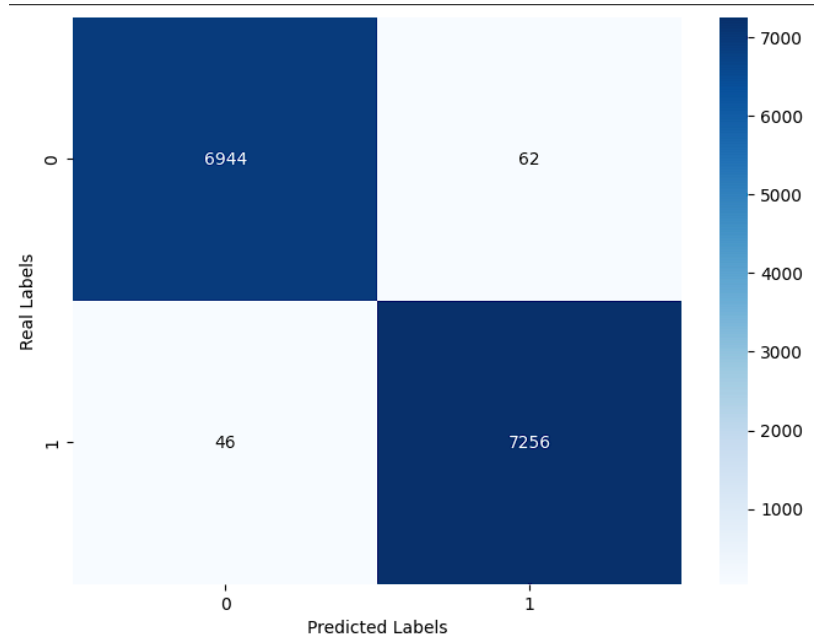
<https://github.com/viferraro/RedesFakeNews/blob/main/WelFakeBertTeste.ipynb>

- Características principais da rede:
  - Rede do tipo Model, sem muitas alterações em relação à anterior.
- Total de Parâmetros: 109.531.521
- Batch-size: 16
- Epochs: 10
- Utiliza um conjunto de validação separado para validar o desempenho do modelo durante o treino.

O treino foi interrompido após 6 épocas, pela configuração do método EarlyStopping. O modelo teve um aparente bom desempenho, com pouca diferença nos valores de treino e validação.



A matriz de confusão também sugere um desempenho muito bom, com 6.944 verdadeiros negativos, 46 falsos negativos, 7.256 verdadeiros positivos e 62 falsos positivos.



Resumo do desempenho:

	precision	recall	f1-score	support
0	0.99	1.00	0.99	7006
1	1.00	0.99	0.99	7302
accuracy			0.99	14308
macro avg	0.99	0.99	0.99	14308
weighted avg	0.99	0.99	0.99	14308

#### 4. Conclusões

Neste trabalho, testei dois modelos de processamento de linguagem natural (PLN) para classificar notícias como verdadeiras ou falsas. Os modelos utilizados foram redes neurais recorrentes (RNNs) com células de longo curto prazo (LSTMs) e transformadores (Bert). Os testes envolveram a modificação de diferentes métodos e hiper parâmetros dos modelos como o tamanho das camadas, dos vocabulários, dos vetores de entrada, dos números de tokens, métodos para introduzir BatchNormalization e regularização dos pesos e bias.

Os resultados mostraram que ambos os modelos tiveram um desempenho muito bom na tarefa de classificação, com exceção de um modelo que apresentou comportamento condizente com provável overfitting (3.2.1).

Mais testes poderiam avaliar melhor o desempenho desses modelos em conjuntos de dados mais desafiadores e com maior número de épocas.

Para um próximo trabalho, testaria redes menos robustas e utilizaria conjuntos de dados mais complexos. Além disso, poderia aumentar o número de épocas de treinamento para avaliar melhor o desempenho dos modelos em longo prazo.

De qualquer forma, os resultados indicam que os modelos de PLN podem ser utilizados para classificar notícias como verdadeiras ou falsas com alta precisão.

## 5. Referências

<https://www.deeplearningbook.com.br/modelo-bert-para-processamento-de-linguagem-natural/>

<https://www.deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory/>

Long Short-Term Memory: [https://en.wikipedia.org/wiki/Long\\_short-term\\_memory](https://en.wikipedia.org/wiki/Long_short-term_memory)