

Análise do Classificador Ingênuo de Bayes para Detecção de Fake News

Eliezer Martins de Oliveira

Erick Vinicius Rebouças Cruz

Joana D'arc Oliveira do

Nascimento

Paulo Vitor Barbosa Santana

Introdução

A proliferação de fake news nas plataformas digitais representa um desafio global para a integridade da informação. Este trabalho avalia o classificador Naive Bayes, método probabilístico eficiente em problemas de alta dimensionalidade típicos de processamento de linguagem natural.

Objetivo

- Avaliar a performance do Naive Bayes na classificação de textos como fake news
- Comparar variantes do algoritmo (MultinomialNB, ComplementNB) em diferentes representações textuais
- Analisar métricas de desempenho como precisão, recall e F1-score

Metodologia

- Abordar a detecção de fake news utilizando técnicas de aprendizado de máquina e análise exploratória de dados textuais.
- Realizar uma análise para compreender a distribuição de termos linguísticos, frequência de palavras-chave e padrões lexicais característicos em notícias falsas.
- Código será desenvolvido em Python, empregando as bibliotecas
 - Scikit-learn para modelagem e vetorização textual
 - Pandas para manipulação de dados
 - NLTK para pré-processamento linguístico
 - Matplotlib e Seaborn para visualização dos padrões textuais e resultados de classificação

Datase

t

O dataset Fake and Real News contém:

- 44,898 artigos
 - 23,481 fake
 - 21,417 reais
- Estrutura típica:
 - Título, conteúdo textual, assunto e rótulo
- Desafios:
 - Variabilidade linguística
 - desbalanceamento de classes

Teorema de Bayes

- Desenvolvido pelo matemático britânico Thomas Bayes (1702–1761)
- É uma das ferramentas da probabilidade, ele oferece um caminho preciso para estimar a chance de ocorrência de fenômenos a partir de conhecimento pré-existente.
- Amplamente reconhecido devido à simplicidade conceitual, aliada a um potencial de uso que atravessa fronteiras disciplinares.
- No aprendizado de máquina e na inteligência artificial, é frequentemente utilizado para desenvolver classificadores probabilísticos

Teorema de Bayes

Matematicamente, o Teorema de Bayes é expresso pela seguinte fórmula:

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

- $P(A|B)$ é a probabilidade do evento A ocorrer, dado o evento B ocorreu.
- $P(B|A)$ é a probabilidade de evento B ocorrer, dado o evento A ocorreu.
- $P(A)$ e $P(B)$ são as probabilidades a priori de A e B, respectivamente.

Classificador Naive Bayes



- Eficaz para processamento de linguagem natural devido à sua capacidade de lidar com alta dimensionalidade de dados textuais.
- Calcula a probabilidade de uma instância pertencer a uma classe específica do conjunto de dados com base em suas características.
- Implementação fácil e resultados precisos em cenários variados
- Aplicações em diversas áreas, como: análise de crédito, análise de texto, campo médico e na engenharia.

Classificador Multinomial Naive Bayes

- Modelo probabilístico amplamente utilizado para a classificação de dados discretos, especialmente em tarefas de processamento de linguagem natural (PLN) e mineração de texto.
- Amplamente empregado em tarefas como classificação de documentos, filtragem de spam e detecção de fake news.
- Apresenta bons resultados, especialmente quando os dados possuem uma distribuição multinomial e as características representam contagens discretas.



Classificador Multinomial Naive Bayes

- Matematicamente, a probabilidade de uma classe dado um vetor de características é calculada como:

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{P(x)}.$$

- $P(x|C_k)$ pode ser expressa como o produto das probabilidades individuais.

$$P(x|C_k) = \prod_{i=1}^n P(x_i|C_k).$$



Classificador Multinomial Naive Bayes

- A probabilidade condicional de uma característica dada a classe pode ser estimada por:

$$P(x_i|C_k) = \frac{\text{count}(x_i, C_k) + \alpha}{\sum_j \text{count}(x_j, C_k) + \alpha N},$$

- Na fase de inferência, um novo exemplo é classificado calculando a pontuação para cada classe e escolhendo aquela com maior valor:

$$\hat{C} = \arg \max_{C_k} \left(\log P(C_k) + \sum_{i=1}^n \log P(x_i|C_k) \right)$$



Classificador Complement



Naive Bayes

- Projetada para lidar com conjuntos de dados desbalanceados, onde certas classes possuem significativamente menos exemplos que outras.
- Eficaz em tarefas como identificação de documentos raros, diagnóstico de anomalias ou categorização hierárquica de textos, onde a abordagem complementar reduz a influência de classes dominantes e melhora a sensibilidade a padrões minoritários.
- O CNB mantém a suposição de independência condicional entre características, mas calcula as probabilidades de forma complementar.



Classificador Complement Naive Bayes

- Matematicamente, a probabilidade de uma classe C_k dado um vetor de características x segue o Teorema de Bayes:

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{P(x)}.$$

- Não estima $P(x|C_k)$ diretamente, o CNB calcula a probabilidade das características no complemento de C_k .

$$P(x|\tilde{C}_k) = \prod_{i=1}^n P(x_i|\tilde{C}_k),$$



Classificador Complement

Naive Bayes

A estimativa da probabilidade condicional complementar para uma característica x_i é suavizada com Laplace e definida como:

$$P(x_i|\tilde{C}_k) = \frac{\sum_{j \neq k} \text{count}(x_i, C_j) + \alpha}{\sum_{j \neq k} \sum_m \text{count}(x_m, C_j) + \alpha N}$$

Na fase de classificação, um novo exemplo é atribuído à classe que minimiza a evidência das características nas classes complementares:

$$\hat{C} = \arg \min_{C_k} \left(\log P(C_k) - \sum_{i=1}^n \log P(x_i|\tilde{C}_k) \right)$$



Protocolo Experimental

- Divisão 80-20 para treino-teste
 - 80% dos dados são utilizados para treino
 - 20% são utilizados para teste
- Validação cruzada estratificada ($k = 5$)
 - Técnica usada para avaliar modelos de aprendizado de máquina, garantindo que cada divisão dos dados preserve a mesma proporção das classes da variável-alvo.
- Matriz de confusão;
- AUC-ROC



Matriz de confusão

- Tabela usada para avaliar o desempenho de um modelo de classificação. Ela compara os valores previstos pelo modelo com os valores reais do conjunto de dados de teste.

	Previsto Positivo	Previsto Negativo
Real Positivo	<i>VP</i>	<i>FN</i>
Real Negativo	<i>FP</i>	<i>VN</i>

- A partir dela calculamos diversas métricas importantes:
 - Acurácia
 - Precisão
 - Recall (sensibilidade)
 - F1-Score



Matriz de confusão

- Acurácia

$$\frac{VP + VN}{VP + VN + FP + FN}$$

- Recall (sensibilidade)

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

- Precisão

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

- F1-Score

$$2 \times \frac{Precisao \times Recall}{Precisao + Recall}$$



AUC-ROC

- A curva ROC é um gráfico que mostra a relação entre:
 - Recall
 - Taxa de Falsos Positivos: Mede a proporção de exemplos negativos que foram classificados incorretamente como positivos.

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

- A AUC-ROC é um número entre 0 e 1 que representa a área sob a curva ROC. Ela indica a capacidade do modelo em distinguir entre as classes.

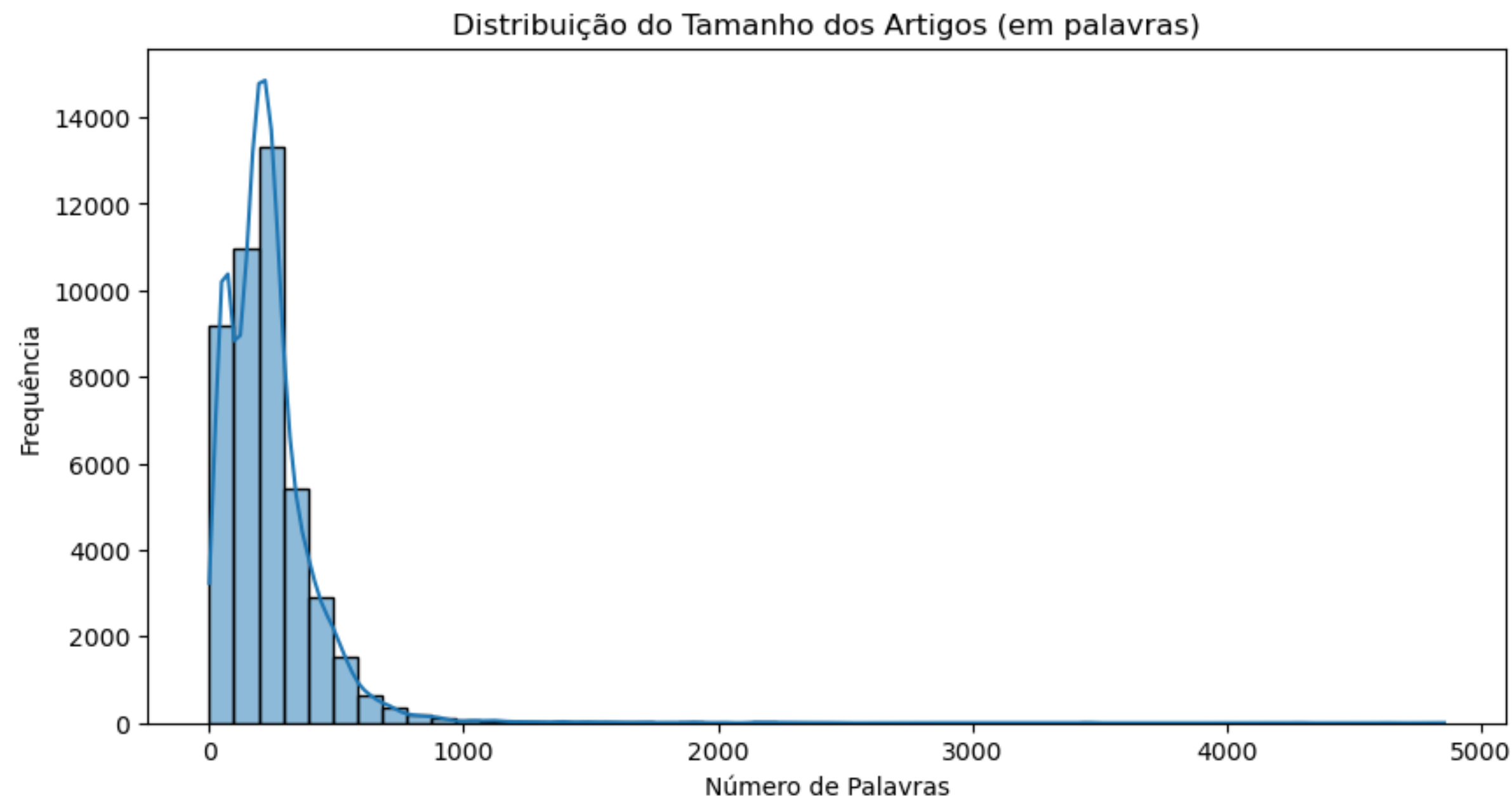


Análise Exploratória de Dados

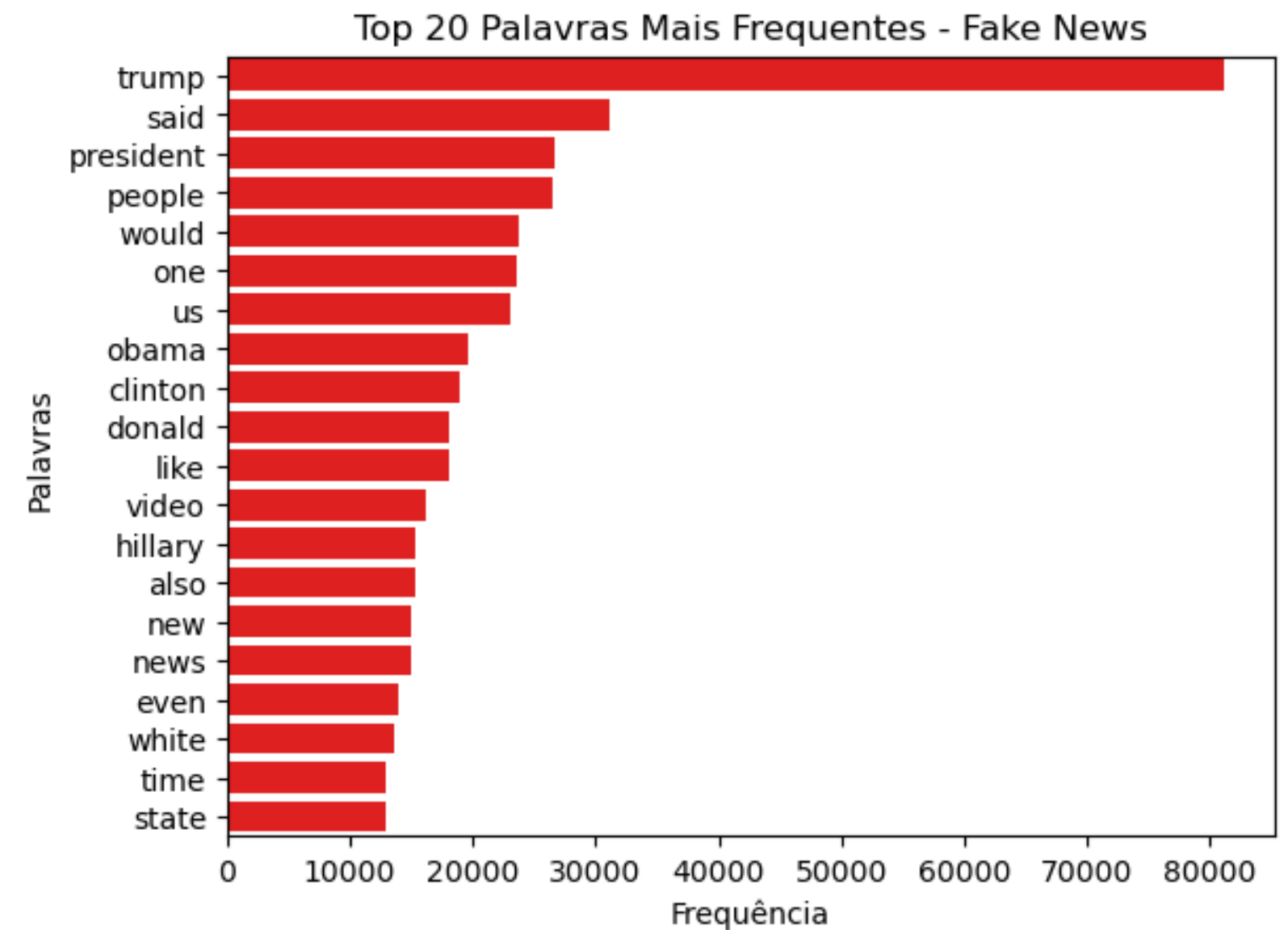
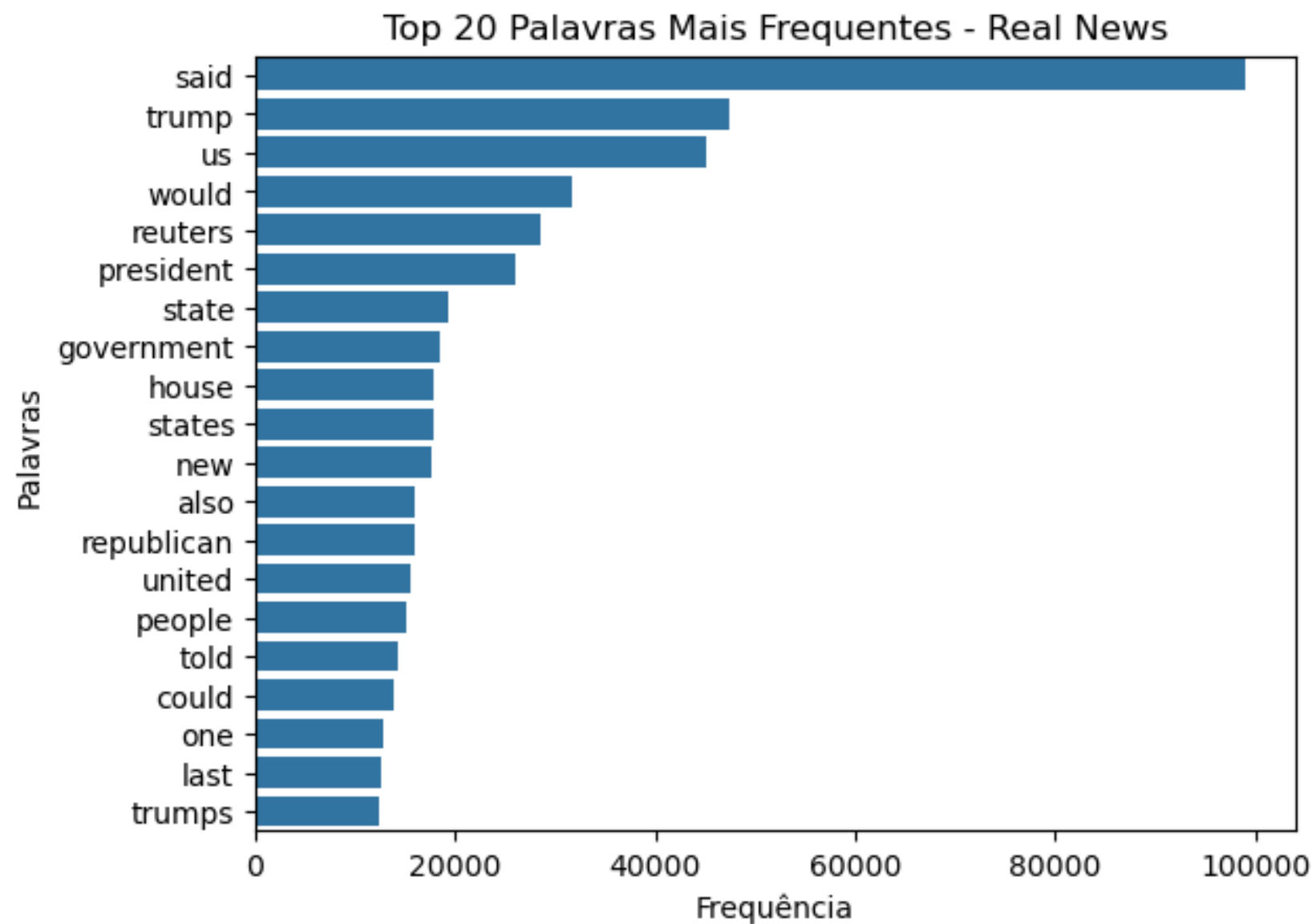
- Garante uma maior qualidade e confiabilidade para o modelo assim sintetizado.
- Permite extrair informações cruciais, tais como a media, a mediana, o desvio padrão, os valores mínimo e máximo, bem como a frequência dos valores associados a uma variável específica.
- Dados Analisados:
 - Tamanho dos Artigos
 - Top 20 Palavras Mais Frequentes
 - Frequência Relativa das Palavras
 - Tamanho das Notícias separado por veracidade

Tamanho dos Artigos

- Distribuição assimétrica
- Maioria dos artigos são curtos, com menos de 300 palavras
- Alguns artigos extensos, na faixa de 4000-5000 palavras



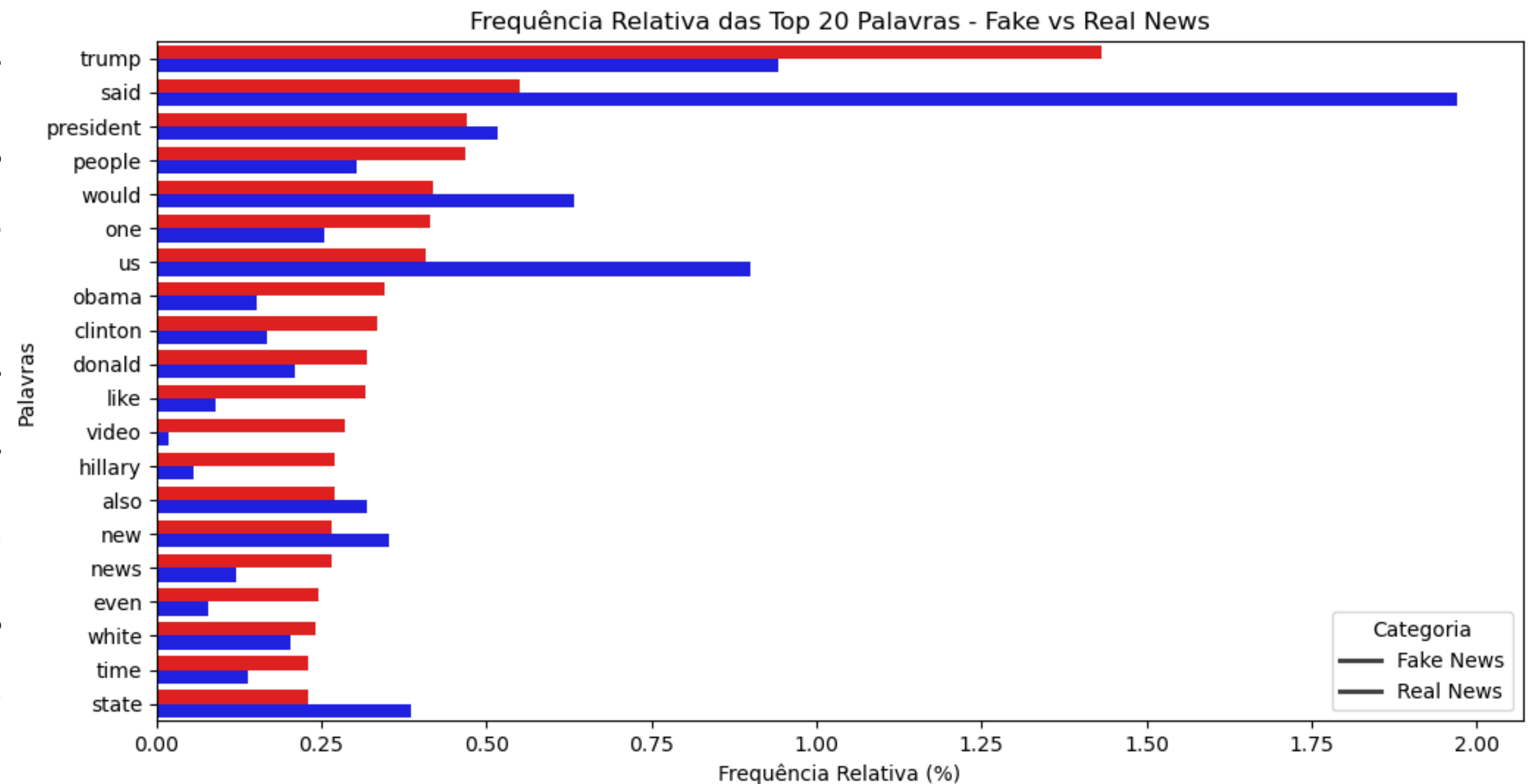
Palavras mais frequentes



- Convergências: Palavras como “trump”, “said” e “president”
- Diferenças: “reuters”, “government” e “house” são mais proeminente em notícias reais, enquanto em notícias falsas surgem nomes de políticos (“obama”, “clinton”, “donald”, “hillary”)

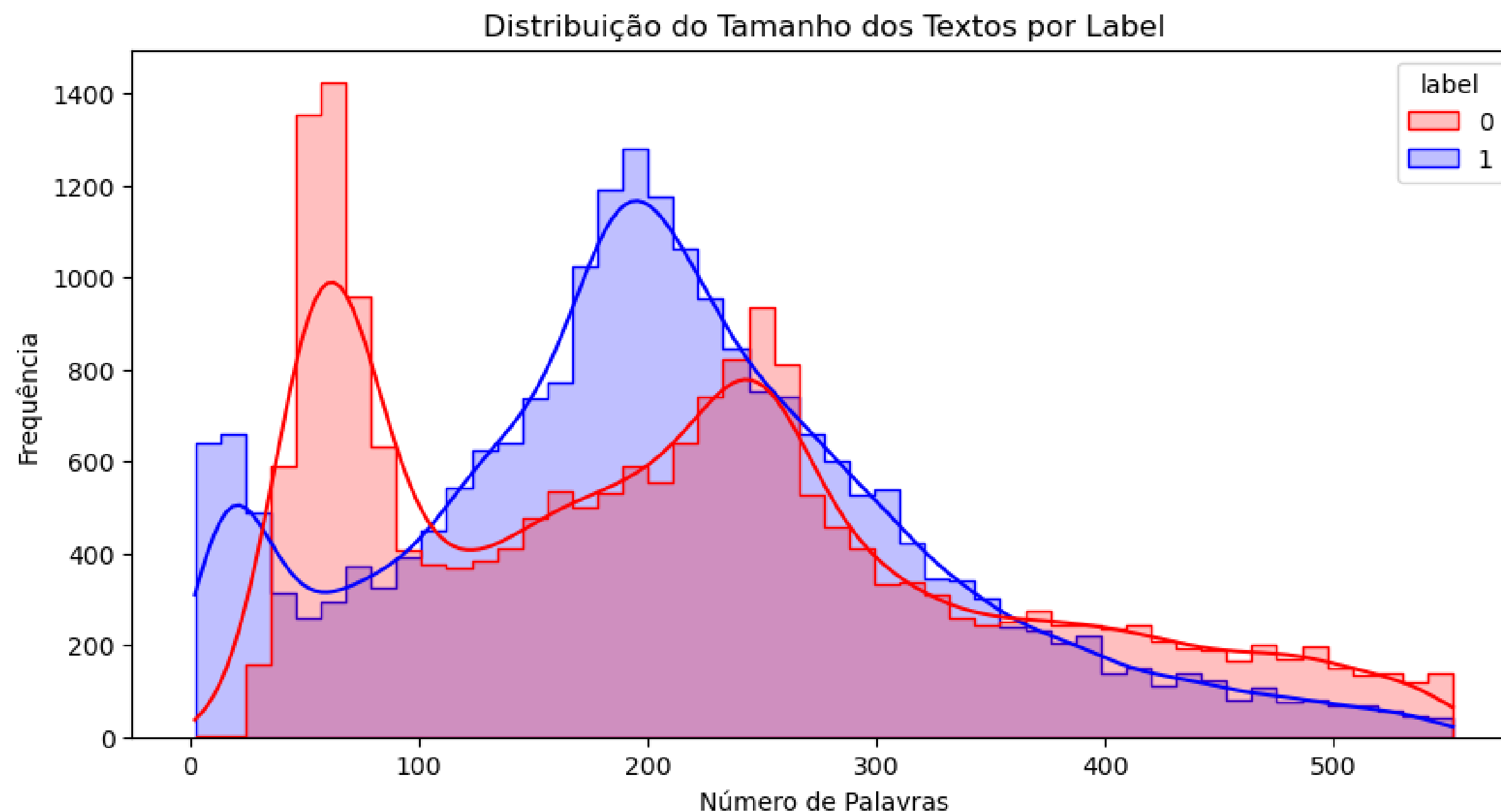
Frequência Relativa das Palavras

Ajuda a identificar padrões característicos em diferentes tipos de textos, dessa forma conseguimos capturar padrões e estruturar dados textuais de forma que algoritmos possam classificá-los de maneira eficiente.



Tamanho das Notícias separado por veracidade

- Fake news mais frequentemente tem menos que 100 palavras
- Real news mais frequentemente tem 200 palavras



Resultado

- Multinomial Naive Bayes

```
Relatório de Classificação:
      precision    recall  f1-score   support

     0       0.95      0.96      0.96      4228
     1       0.97      0.95      0.96      4752

 accuracy              0.96      8980
 macro avg           0.96      0.96      0.96      8980
 weighted avg       0.96      0.96      0.96      8980
```

```
Acurácia: 0.9589086859688196
Precision (MultinomialNB): 0.9673704414587332
AUC-ROC: 0.9836283642161127
AUC-PR: 0.9860810878125684
```


Resultado

- Complement Naive Bayes

```
Relatório de Classificação:
      precision    recall  f1-score   support

     0       0.95      0.96      0.96      4228
     1       0.97      0.95      0.96      4752

 accuracy          0.96          8980
 macro avg       0.96      0.96      0.96      8980
weighted avg       0.96      0.96      0.96      8980
```

```
Acurácia: 0.9586859688195991
Precision (ComplementNB): 0.9673565180285897
AUC-ROC: 0.9836275180852996
AUC-PR: 0.9860801940006345
```

Resultado

- Ambos modelos altamente eficazes, com valores parecidos
- Acurácia está em torno de 95.9%
- Quando o modelo prevê uma notícia falsa, ele está certo em aproximadamente 96.7% das vezes
- A AUC-ROC é aproximadamente 98%
- Os modelos conseguem distinguir bem entre notícias falsas e verdadeiras

Conclusão

- Os resultados mostram que ambos os modelos se mostraram altamente eficazes na classificação de Fake News, com desempenho quase idêntico em todas as métricas avaliadas.
- Os resultados indicam que o Naive Bayes continua sendo uma excelente abordagem para classificação automática de Fake News, devido à sua simplicidade, eficiência e alto desempenho.
- Para aprimorar ainda mais os resultados, podem ser exploradas técnicas como engenharia de características, TF-IDF para vetorização de texto, ou até mesmo a combinação com outros classificadores mais complexos.

Fim

!

