



Total de amostras: 18896
Notícias falsas: 9480
Notícias verdadeiras: 9416

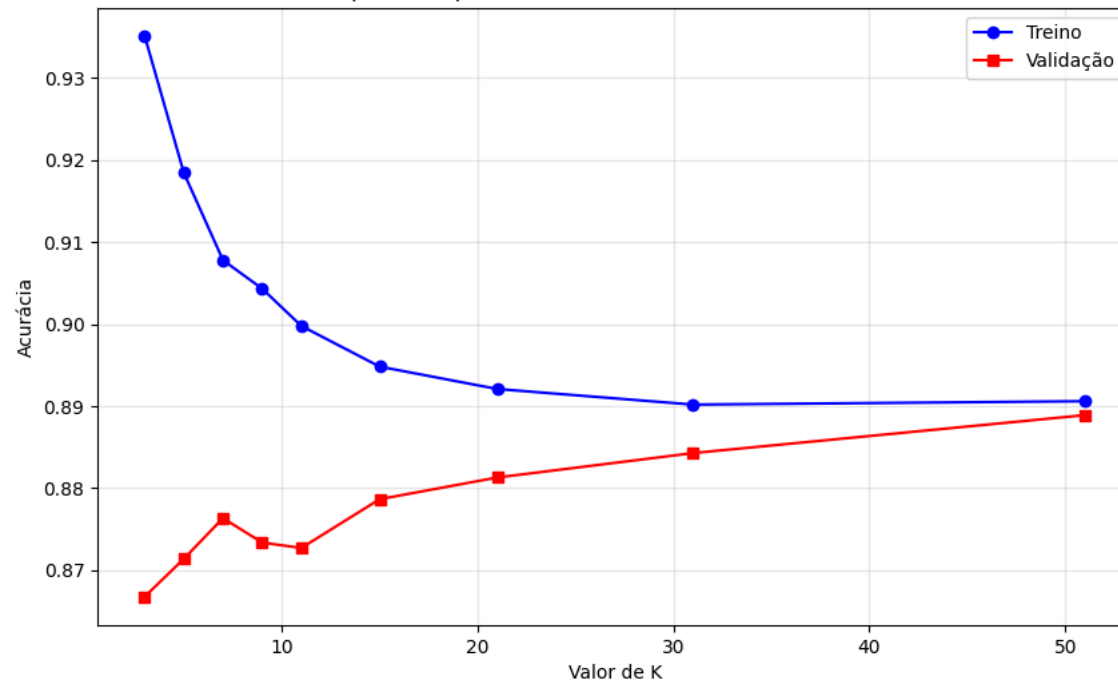
Shape da matriz TF-IDF: (18896, 5000)
Número de features (palavras/termos): 5000

Tamanho do conjunto de treino: 15116
Tamanho do conjunto de teste: 3780

===== ANÁLISE DE DIFERENTES VALORES DE K =====

K=3: Acurácia Treino=0.9352, Acurácia Validação=0.8667
K=5: Acurácia Treino=0.9185, Acurácia Validação=0.8714
K=7: Acurácia Treino=0.9078, Acurácia Validação=0.8763
K=9: Acurácia Treino=0.9043, Acurácia Validação=0.8733
K=11: Acurácia Treino=0.8998, Acurácia Validação=0.8727
K=15: Acurácia Treino=0.8948, Acurácia Validação=0.8786
K=21: Acurácia Treino=0.8921, Acurácia Validação=0.8813
K=31: Acurácia Treino=0.8902, Acurácia Validação=0.8843
K=51: Acurácia Treino=0.8906, Acurácia Validação=0.8889

Impacto do parâmetro K na acurácia do modelo KNN



Melhor valor de K baseado na validação: 51

===== VALIDAÇÃO CRUZADA =====

Scores da validação cruzada: [0.88988095 0.88223619 0.88190539 0.89348329 0.87992061]
Acurácia média (CV): 0.8855 (+/- 0.0105)

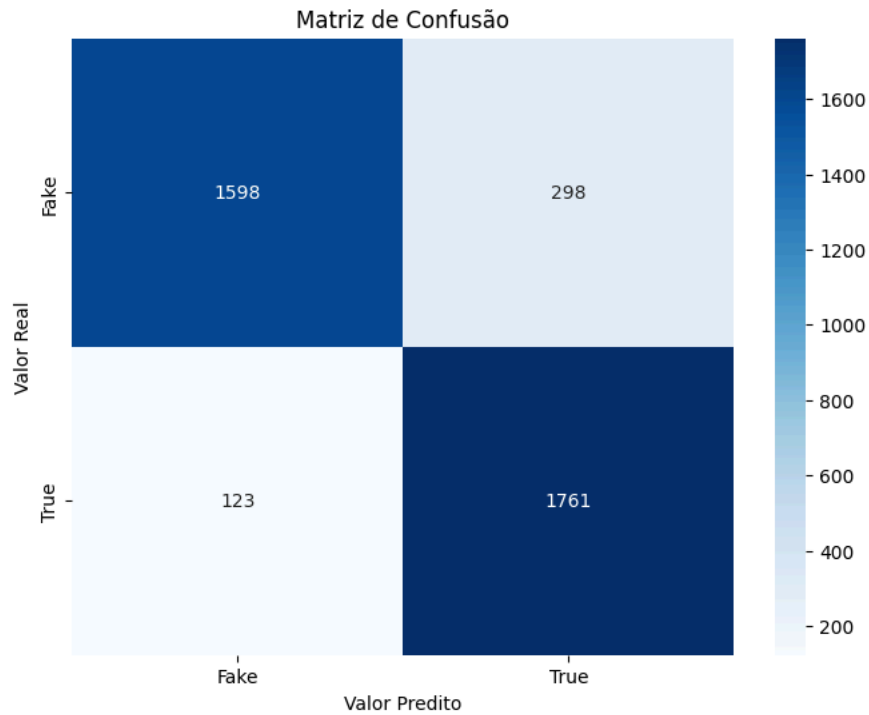
Resultados detalhados da validação cruzada:

Accuracy: 0.8855 (+/- 0.0105)

Precision: 0.8530 (+/- 0.0151)

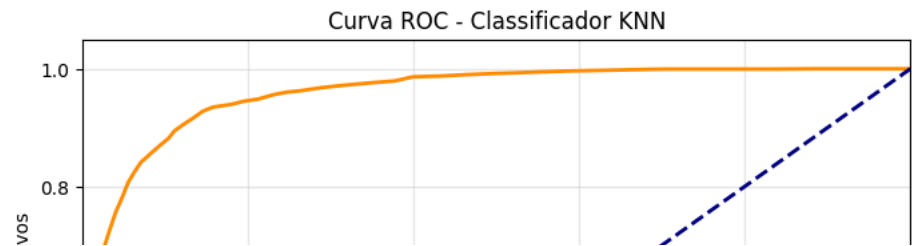
Recall: 0.9307 (+/- 0.0140)
F1: 0.8901 (+/- 0.0097)

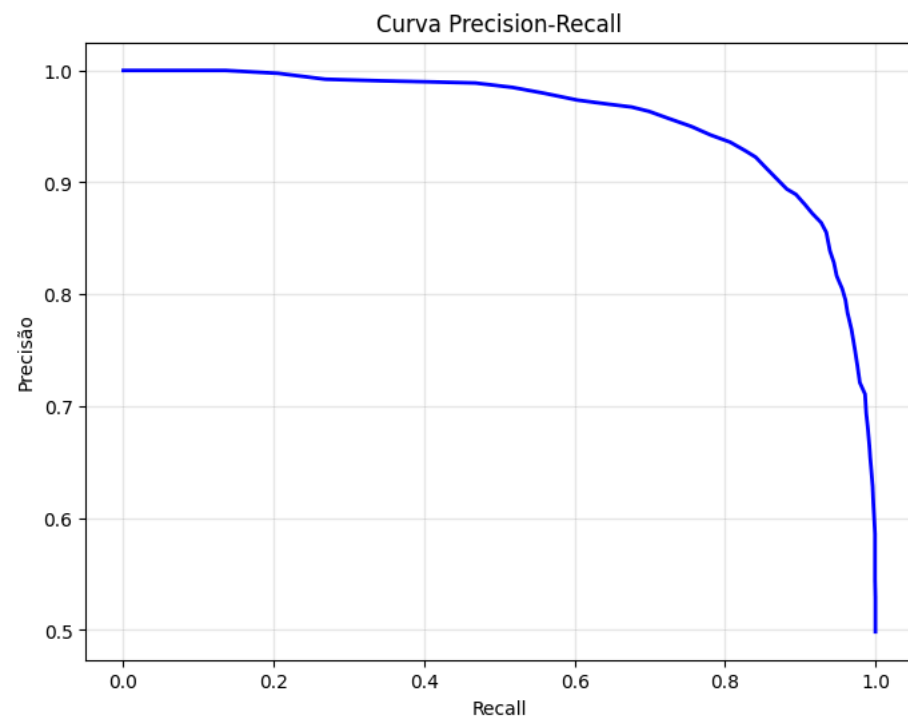
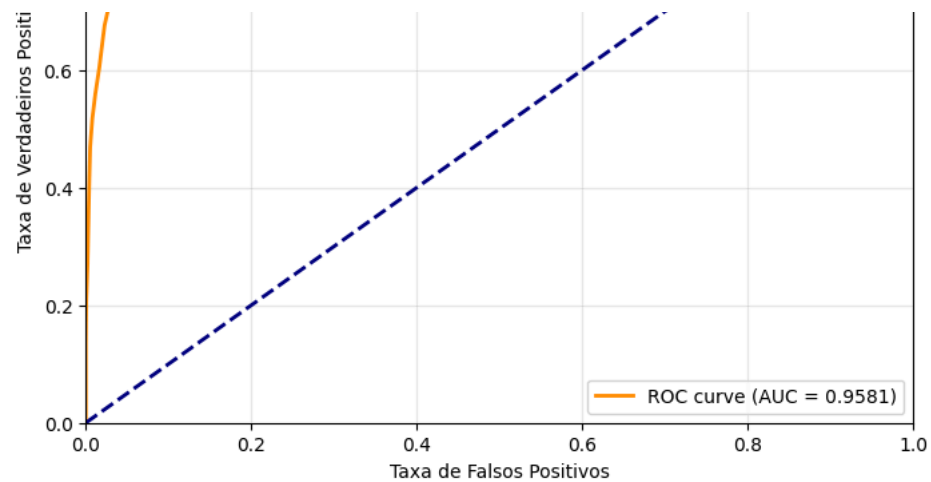
===== AVALIAÇÃO DO CLASSIFICADOR BINÁRIO =====
Acurácia: 0.8886
Precisão: 0.8553
Recall (Sensibilidade): 0.9347
F1-Score: 0.8932
Especificidade: 0.8428



Relatório de Classificação:

	precision	recall	f1-score	support
Fake News	0.93	0.84	0.88	1896
True News	0.86	0.93	0.89	1884
accuracy			0.89	3780
macro avg	0.89	0.89	0.89	3780
weighted avg	0.89	0.89	0.89	3780





===== INTERPRETAÇÃO DOS RESULTADOS =====

ANÁLISE DA EFICIÊNCIA DO CLASSIFICADOR:

1. ACURÁCIA (0.8886):
 - O modelo classifica corretamente 88.86% das notícias
 - Indica boa performance geral na separação entre notícias falsas e verdadeiras
2. PRECISÃO (0.8553):

- Quando o modelo classifica uma notícia como verdadeira, está correto em 85.53% dos casos
- Alta precisão minimiza falsos positivos (notícias falsas classificadas como verdadeiras)

3. RECALL/SENSIBILIDADE (0.9347):

- O modelo identifica 93.47% de todas as notícias verdadeiras
- Importante para não deixar passar notícias verdadeiras como falsas

4. ESPECIFICIDADE (0.8428):

- O modelo identifica corretamente 84.28% das notícias falsas
- Complementa o recall, mostrando a capacidade de detectar negativos

5. F1-SCORE (0.8932):

- Média harmônica entre precisão e recall
- Valor de 0.8932 indica muito bom equilíbrio entre precisão e recall

6. AUC-ROC (0.9581):

- Área sob a curva ROC próxima de 1.0 indica excelente capacidade discriminativa
- O modelo tem excelente capacidade de distinguir entre classes

CONCLUSÃO:

O classificador KNN com TF-IDF demonstra boa eficiência na detecção de notícias falsas. O valor de K=51 proporciona o melhor equilíbrio entre complexidade e generalização. As métricas indicam que o modelo é confiável para uso prático, com poucos erros tanto de falsos positivos quanto falsos negativos.

ANÁLISE DOS ERROS:

Total de erros: 421

Falsos Positivos (Fake classificada como True): 298

Falsos Negativos (True classificada como Fake): 123

Proporção FP/FN: 2.42

