

A etapa de deduplicação de dados, com foco especial na integração de conjuntos de dados de produtos da Amazon e do Google. Com um tempo de execução de 11.74 segundos, o sistema demonstrou ser relativamente rápido e eficiente, levando em consideração a complexidade e o volume dos dados processados.

A precisão alcançada de 87,12% indica que a maioria das duplicatas identificadas pelo algoritmo são realmente duplicatas, sugerindo que o sistema é altamente confiável e minimiza os falsos positivos. O recall de 94,64% revela que o algoritmo conseguiu identificar uma grande proporção das duplicatas reais presentes nos conjuntos de dados, garantindo que poucas duplicatas genuínas tenham sido negligenciadas. O F1-Score de 90,73% fornece uma medida balanceada do desempenho geral do sistema, refletindo uma excelente harmonia entre precisão e recall. Essas métricas não apenas validam a eficácia do processo de deduplicação, mas também destacam sua capacidade de fornecer uma base de dados limpa e confiável para análises subsequentes e tomadas de decisão.

Com base nesses resultados, podemos concluir que a metodologia proposta para deduplicação de dados é robusta e eficaz, sendo adequada para o cenário proposto em nossa abordagem. O sistema se mostrou uma ferramenta valiosa para a limpeza e preparação de dados, elementos essenciais para qualquer análise de dados de alta qualidade.

```
Python 3.7.8 (tags/v3.7.8:4b47a5b6ba, Jun 28 2020, 08:53:46) [MSC v.1916 64 bit  
(AMD64)] on win32  
Type "help", "copyright", "credits" or "license()" for more information.  
>>>  
|===== RESTART: C:\Users\Gabriel\Documents\DEDUPLICAÇÃO_BDA\algoritmo.py =====  
Precision: 0.8712  
Recall: 0.9464  
F1-Score: 0.9073  
Tempo de execução do algoritmo: 11.74 segundos.  
>>>
```