**Tìm hiểu về thuật toán Apriori**

Apriori sử dụng kiến thức đã có từ trước (prior) về các thuộc tính, vật phẩm thường xuyên xuất hiện trong cơ sở dữ liệu. Để cải thiện hiệu quả của việc lọc các mục thường xuyên theo cấp độ, một thuộc tính quan trọng được sử dụng gọi là thuộc tính Apriori giúp giảm phạm vi tìm kiếm của thuật toán.

**1.Thuật toán**

Thuật toán Apriori hoạt động theo cơ chế "đóng gói" mục (itemset) từ các subset của chúng. Tội bán cần tìm ra tất cả các itemset phổ biến trong dữ liệu tập hợp lớn, tức là tập hợp các itemset xuất hiện trong dữ liệu nhiều hơn hoặc bằng một độ xuất hiện được xác định trước đó.

Các bước thực hiện của thuật toán Apriori gồm:

1. Tìm tất cả các mục đơn lẻ (single item).

2. Tạo một bảng tần suất cho các mục đơn lẻ.

3. Dựa trên bảng tần suất, đánh giá tất cả các itemset bao gồm ít nhất hai mục.

4. Loại bỏ itemset có tần suất phổ biến thấp hơn một ngưỡng được xác định trước để giảm bớt chi phí xử lý.

5. Tiếp tục quá trình đóng gói cho đến khi không còn itemset nào phổ biến hơn.

+) Các bước 2 đến 5 được lặp lại cho đến khi không còn itemset nào phổ biến hơn. Sau đó, thuật toán Apriori trả về tất cả các itemset phổ biến được tìm thấy.

+) Trong quá trình thực hiện, thuật toán Apriori sử dụng một số kỹ thuật tối ưu như “pruning” để cắt bớt các nhánh không có thể trở thành các tập hợp phổ biến, giúp cải thiện tốc độ xử lý của thuật toán.

**2.Ứng dụng của thuật toán Apriori**

**Ứng dụng**

Apriori được sử dụng để khai thác dữ liệu để tìm ra các mục xuất hiện cùng nhau trong các tập dữ liệu. Thuật toán này có thể được sử dụng để phân tích hành vi của khách hàng, gợi ý sản phẩm và dịch vụ cho khách hàng, và xác định các chuỗi sản phẩm của cửa hàng, giảm thiểu lãng phí, tối ưu hóa kinh doanh và nhiều ứng dụng khác.

**Ưu điểm**

* Nó được sử dụng để tính toán các tập phổ biến lớn.
* Đơn giản để hiểu và áp dụng.

**Nhược điểm**

* Apriori yêu cầu lưu trữ toàn bộ tập dữ liệu trong bộ nhớ để tính toán các tập phổ biến.
* Độ phức tạp tính toán cao: Quá trình tìm kiếm các tập phổ biến có thể yêu cầu nhiều thời gian và tài nguyên tính toán.
* Không xử lý được dữ liệu lớn: Khi dữ liệu tăng lên, Apriori có thể không xử lý được các tập dữ liệu lớn.

**3.Bài toán thực tế**

Cửa hàng làm một combo gồm ngũ cốc, sữa và bánh mì. Cửa hàng cũng giảm giá cho những khách hàng của họ mua những combo này. Bằng cách làm combo, cửa hàng đã làm cho khách hàng dễ dàng lựa chọn các sản phẩm. Đồng thời cũng làm tăng hiệu suất bán hàng của mình.

Cửa hàng đã sử dụng thuật toán apriori để phân tích hành vi của khách hàng.

Cụ thể, Apriori hoạt động bằng cách tìm kiếm các mặt hàng thường được mua hàng đầu và dựa trên đó, tìm kiếm các giá trị còn lại trong tập dữ liệu. Sử dụng thuật toán FP để tìm kiếm các mặt hàng, và sau đó sử dụng một quá trình quét để tìm kiếm các luật kết hợp.

Apriori có thể giúp doanh nghiệp phát hiện các sản phẩm phổ biến được bán kèm với nhau, từ đó tăng doanh số bằng cách đề xuất các gợi ý mua hàng.

Nó giúp khách hàng mua sản phẩm của họ một cách dễ dàng và tăng hiệu suất bán hàng, đồng thời giúp cửa hàng nắm bắt và điều chỉnh các sản phẩm để có thể phù hợp với yêu cầu mua hàng của khách hàng.