**Exercise 1: Online Extraction vs Offline Extraction. Example?**

Online Extraction và Offline Extraction là hai phương pháp được sử dụng trong khai phá dữ liệu và truy xuất thông tin để lấy thông tin từ các nguồn khác nhau.

**Online Extraction**

**Định nghĩa:** Online Extraction bao gồm việc trích xuất thông tin từ luồng dữ liệu hoặc nguồn dữ liệu thời gian thực.

**Đặc điểm:**

* Quá trình liên tục: Online Extraction được thực hiện liên tục khi dữ liệu mới trở nên có sẵn.
* Xử lý ngay lập tức: Dữ liệu được trích xuất và xử lý theo thời gian thực hoặc gần thời gian thực.

Ví dụ: Theo dõi luồng cập nhật mạng xã hội cho các từ khóa hoặc hashtag cụ thể để theo dõi xu hướng hoặc tâm trạng. Khi các bài đăng mới được đăng, hệ thống trích xuất thông tin liên quan trong thời gian thực.

**Offline Extraction**

**Định nghĩa**: Offline Extractoin bao gồm trích xuất thông tin từ các nguồn dữ liệu tĩnh không thay đổi thường xuyên.

**Đặc điểm:**

* Xử lý theo lô: Dữ liệu được trích xuất theo lô, và quá trình trích xuất không liên tục.
* Dữ liệu lịch sử: Thường được sử dụng để trích xuất thông tin từ dữ liệu lịch sử hoặc dữ liệu được lưu trữ.

**Ví dụ:** Phân tích đánh giá của khách hàng từ một tập dữ liệu được thu thập trong vài năm để xác định phàn nàn phổ biến hoặc xu hướng. Quá trình trích xuất không cần thiết phải theo thời gian và có thể được thực hiện hàng loạt.

**Sự khác biệt chính:**

* Thời gian: Online Extraction được thực hiện trong thời gian thực hoặc gần thời gian thực, trong khi Offline Extraction được thực hiện trên dữ liệu tĩnh.
* Nguồn dữ liệu: Online Extraction xử lý các nguồn dữ liệu động, luồng, trong khi Offline Extraction xử lý các nguồn dữ liệu tĩnh, lịch sử.
* Phương pháp xử lý: Online Extraction đòi hỏi xử lý ngay lập tức, thường sử dụng thuật toán luồng, trong khi Offline Extraction có thể sử dụng các thuật toán xử lý theo lô.