

Mô hình CRISP-DM và các giai đoạn của qui trình

CRISP – DM (Cross Industry Standard Process for Data Mining) là quy trình phân tích dữ liệu tiêu chuẩn, giúp đảm bảo các dự án phân tích dữ liệu có tính hệ thống, giảm rủi ro và dễ tái sử dụng được hình thành từ năm 1996. Mô hình CRISP-DM gồm **6 giai đoạn chính**:

1. Hiểu về nhu cầu của doanh nghiệp (Business Understanding):

Bước này về cơ bản là để xác định mục tiêu kinh doanh (business objective) của dự án và đánh giá tiềm năng của các giải pháp khoa học dữ liệu.

2. Hiểu về dữ liệu (Data Understanding):

Thu thập dữ liệu ban đầu, kiểm tra chất lượng, hiểu cấu trúc, nội dung dữ liệu và đánh giá xem dữ liệu có phù hợp cho các giải pháp đề ra hay không.

3. Chuẩn bị dữ liệu (Data Preparation):

Sau khi xác định nguồn, dữ liệu cần được chọn lọc, làm sạch, chuyển đổi và tích hợp để tạo thành bộ dữ liệu gọn, rõ ràng, phù hợp cho phân tích và tránh nhầm lẫn.

4. Xây dựng mô hình (Modeling):

Áp dụng các thuật toán lên dữ liệu (được chuẩn bị ở bước trước đó). Do mỗi thuật toán có ưu nhược điểm khác nhau nên ta sẽ xét theo các tiêu chí: **độ chính xác, khả năng diễn giải, thời gian huấn luyện mô hình**.

5. Đánh giá chất lượng mô hình (Evaluation):

Đánh giá chất lượng mô hình bằng cách phân tích, cải tiến và so sánh kết quả thu được với mục tiêu nghiên cứu và kinh doanh ban đầu, nhằm xác định mức độ hữu ích, tính khả thi và đóng góp của mô hình vào chiến lược kinh doanh.

6. Triển khai giải pháp (Deployment):

Giai đoạn này nhằm chuyển kết quả phân tích thành chiến lược cụ thể và trình bày bằng công cụ trực quan như biểu đồ, báo cáo để nhà quản lý dễ hiểu và trao đổi với chuyên gia.

Tài liệu tham khảo:

<https://www.studocu.vn/vn/document/truong-dai-hoc-kinh-te-thanh-pho-ho-chi-minh/khoa-hoc-du-lieu/mo-hinh-crisp-bai-tap-nghien-cuu-mo-hinh/122485342>

<https://www.pandaml.com/quy-trinh-khai-thac-du-lieu-trong-doanh-nghiep/>

