Tên: Huỳnh Nhật Thành

MSSV: 3123580044

**Mô hình CRISP-DM và các giai đoạn**

**CRISP-DM** (Cross Industry Standard Process for Data Mining) là quy trình chuẩn trong khai phá dữ liệu, được áp dụng rộng rãi trong các dự án khoa học dữ liệu và học máy. Mô hình này gồm 6 giai đoạn chính, có tính lặp lại và linh hoạt, giúp dự án đi đúng hướng và đảm bảo giá trị thực tiễn.

1. **Business Understanding (Hiểu biết nghiệp vụ)**  
   Xác định mục tiêu kinh doanh và yêu cầu thực tế. Các câu hỏi nghiệp vụ được chuyển thành mục tiêu phân tích dữ liệu cụ thể.
2. **Data Understanding (Hiểu biết dữ liệu)**  
   Thu thập và khám phá dữ liệu ban đầu. Kiểm tra chất lượng, phát hiện dữ liệu thiếu hoặc bất thường, đồng thời tìm hiểu mối quan hệ sơ bộ giữa các biến.
3. **Data Preparation (Chuẩn bị dữ liệu)**  
   Làm sạch và biến đổi dữ liệu, chọn lọc thuộc tính quan trọng, tích hợp các nguồn dữ liệu. Đây là bước chiếm nhiều thời gian nhất và tạo ra tập dữ liệu “sạch” cho mô hình.
4. **Modeling (Xây dựng mô hình)**  
   Áp dụng các thuật toán thống kê và học máy (ví dụ: hồi quy logistic, cây quyết định, SVM). Thử nghiệm nhiều mô hình và điều chỉnh tham số để đạt kết quả tối ưu.
5. **Evaluation (Đánh giá)**  
   Đánh giá hiệu quả mô hình bằng cả chỉ số kỹ thuật (accuracy, precision, recall) và khả năng đáp ứng mục tiêu kinh doanh ban đầu.
6. **Deployment (Triển khai)**  
   Đưa mô hình vào thực tế, có thể dưới dạng báo cáo hoặc tích hợp vào hệ thống. Cần theo dõi và bảo trì để mô hình duy trì hiệu quả khi dữ liệu thay đổi.

👉 Tóm lại, CRISP-DM là một quy trình chặt chẽ từ hiểu nghiệp vụ, xử lý dữ liệu, xây dựng và triển khai mô hình, giúp đảm bảo kết quả phân tích dữ liệu vừa chính xác vừa hữu ích.